

Estimation statistique de la fréquence des tarés en génétique humaine

THÈSE

présentée à la Faculté des sciences de l'Université de Genève
pour obtenir le grade de Docteur en statistique mathématique

par

Aloïs KAELIN

de

EINSIEDELN (SCHWYZ)

Thèse N° 1270



Digitized by the Internet Archive
in 2026

https://archive.org/details/IA41552322_0003

Estimation statistique de la fréquence des tarés en génétique humaine

THÈSE

présentée à la Faculté des sciences de l'Université de Genève
pour obtenir le grade de Docteur en statistique mathématique

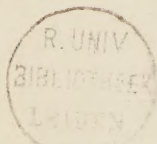
par

Aloïs KAELIN

de

EINSIEDELN (SCHWYZ)

Thèse N° 1270



Editions Médecine et Hygiène
GENÈVE
1957

La faculté des sciences, sur le préavis de Monsieur le professeur Arthur Linder, autorise l'impression de la présente thèse, sans exprimer d'opinion sur les propositions qui y sont énoncées.

Genève, le 10 juillet 1957.

*Le doyen :
Fernand CHODAT*

Thèse N° 1270

Le présent travail, et l'étude des problèmes statistiques en génétique humaine plus généralement, ont été suggérés à l'auteur lorsqu'il travaillait au Laboratoire de Statistique Mathématique de l'Université de Genève (Directeur: Prof. A. Linder) en contact étroit avec l'Institut de Génétique médicale de la Clinique d'Ophtalmologie (Prof. Dr. A. Franceschetti, Dr. D. Klein). L'auteur est redevable à MM. A. Linder, A. Franceschetti et D. Klein des précieux renseignements fournis et des nombreuses remarques faites par eux quant au contenu de cette étude.

L'auteur désire également exprimer sa reconnaissance à Mlle Thérèse Siegrist pour la part considérable qu'elle a prise dans la vérification des tableaux, ainsi qu'à Mlle Rita Linder et M. Richard Lang (Laboratoire de Statistique Mathématique, Genève) pour l'aide qu'ils ont apportée à la préparation du manuscrit; M. Lang s'est en outre chargé de dessiner les graphiques.

TABLE DES MATIERES

- I. INTRODUCTION
- II. QUELQUES NOTIONS FONDAMENTALES DE STATISTIQUE
 - Espérance mathématique du nombre de sujets atteints
 - Variance du nombre r de sujets atteints
 - Espérances mathématiques et variances dans les répartitions tronquées et non tronquées
- III. ESTIMATION DE LA FREQUENCE DE CARACTERE P PAR IDENTIFICATION DES OBSERVATIONS ET DE L'ESPERANCE MATHEMATIQUE. METHODE DU MAXIMUM LIKELIHOOD. METHODE DES SCORES.
 - Méthode d'identification de l'observation et de l'espérance mathématique
 - Le procédé du Maximum Likelihood
 - D-scores
 - Les scores pondérés de FINNEY
- IV. LES PROCEDES D'ESTIMATION DE WEINBERG, RIDER ET MOORE
 - Le procédé de WEINBERG
 - Le procédé de RIDER
 - Le procédé de MOORE
 - Comparaison des différents procédés d'estimation
- V. L'ESTIMATION DE LA FREQUENCE DE CARACTERE P A PARTIR DE FRATRIES DE GRANDEUR ARBITRAIRE
 - Méthode de WEINBERG, RIDER et MOORE
 - Méthode du Maximum Likelihood et méthode des scores
 - Exemples d'application
- RESUME - SUMMARY - ZUSAMMENFASSUNG
- BIBLIOGRAPHIE
- TABLES - TABLEAUX

Travail du Laboratoire de Statistique mathématique de l'Université de Genève
(Directeur: Prof. A. Linder)

ESTIMATION STATISTIQUE DE LA FRÉQUENCE DES TARÉS EN GÉNÉTIQUE HUMAINE

par ALOIS KÄELIN

I. INTRODUCTION

POUR analyser la transmission d'un caractère dont on suppose qu'il est héréditaire, il est nécessaire de connaître entre autres la fréquence relative des porteurs du caractère héréditaire (ou sujets atteints) dans les fratries; il faut également comparer la fréquence observée à celle déduite d'une hypothèse héréditaire donnée, soit l'hérédité mendélienne simple dominante ou récessive, ce qui permet de porter un jugement sur le comportement héréditaire d'un caractère. Dans le cas d'un caractère obéissant aux lois de l'hérédité simple, il n'y a pas, en général, de difficulté particulière à classer, selon leur type héréditaire, un nombre suffisant de fratries en se basant sur la constitution héréditaire de leurs ancêtres, et à établir ainsi une série non sélectionnée de fratries qui permettent l'estimation immédiate de la fréquence de caractère et l'examen de l'hypothèse héréditaire en question.

Les choses se présentent sous un jour moins favorable lorsqu'il s'agit de caractères rares, le plus souvent d'anomalies graves ou d'infirmités, et le cas n'est pas rare où ils excluent même la reproduction de ceux qui en sont atteints. Ainsi, des caractères récessifs se révèlent, en général, lorsque deux parents hétérozygotes et d'apparence normale se reproduisent et ont des descendants homozygotes récessifs chez qui le caractère se manifeste. Une classification des fratries, avec ou sans porteur de caractère, n'est pas possible ici, car deux parents hétérozygotes pour le caractère récessif peuvent avoir des descendants sains seulement. Si l'on doit, par la force des choses, se limiter à des fratries d'un porteur de caractère au moins, on n'obtient pas un matériel non-sélectionné, car les fratries qui ne présentent pas de sujets atteints nous échappent, bien qu'elles pourraient en présenter selon la constitution génétique des parents. En conséquence, lorsqu'on calcule la fréquence relative des porteurs de caractère pour des fratries d'un porteur de caractère au moins, on constate, même pour un caractère récessif vrai, un excès de récessifs. Il en va de même pour des caractères dominants rares, incomplètement pénétrants, du moins s'il s'agit de fratries de parents d'apparence normale; la sélection et l'estimation de la fréquence des porteurs de caractère qui en résulte sont biaisées.

De plus, il est évident que la façon d'examiner les sujets atteints, c'est-à-dire les fratries, a une influence décisive sur le résultat d'une sélection. Si la sélection porte, en premier lieu, sur le sujet atteint pris isolément et ensuite seulement sur la fratrie dont il fait partie, la chance d'une fratrie d'être examinée augmente avec le nombre de sujets atteints qu'elle comporte. Mais si l'examen vise d'abord les fratries, pour autant qu'elles présentent des porteurs de caractère, et qu'on saisit dans une région, toutes les fratries avec porteurs de caractère, ou du moins un échantillon représentatif, l'examen d'une fratrie ne dépendra pas essentiellement du fait qu'un seul ou tous ses membres soient porteurs de caractère. Le procédé mentionné en premier lieu est caractéristique de la sélection par probands, le second est connu sous le nom de sélection par familles.

Pourtant, dans un cas concret, les conditions de sélection sont en général beaucoup plus complexes. Le plus souvent, il s'agit d'une des formes suivantes de sélection par probands : soit qu'un patient entre en clinique et que, par la suite, une anamnèse de famille révèle d'autres porteurs de caractère, soit que les porteurs de caractère d'une fratrie viennent à être observés dans une clinique, indépendamment les uns des autres, soit encore, pour des caractères très rares, que les cas isolés aient été recherchés dans la littérature. Dans la sélection par probands, l'examen des porteurs de caractère est sujet à maints hasards que l'on peut très difficilement saisir et dont la considération est assez problématique : que l'on pense un peu à la gravité variable de la maladie, aux chances de guérison qui peuvent déterminer ou non le malade à consulter un médecin ou même à l'intérêt, obligatoirement limité par le temps, que le médecin peut porter à la maladie. On a tenté, il est vrai, de tenir compte de certaines de ces conditions de sélection en introduisant des paramètres appropriés dans les modèles statistiques, par exemple de tenir compte de la sélection multiple d'une fratrie par plusieurs de ses probands (« indépendants ») en introduisant la probabilité qu'un sujet atteint devienne proband (degré de sélection). Sous ce rapport, il faut également citer la méthode par le premier proband. Mais en général, l'application de ces procédés spéciaux se heurte à l'insuffisance des données sur le nombre des sujets atteints examinés indépendamment, ou sur le mode de sélection lui-même.

C'est pourquoi, par la suite, on ne commentera que les deux cas limites de la méthode par probands : Premier cas : tous les porteurs de caractère sont considérés comme probands (sélection complète, « complete selection »), le degré de sélection ayant la valeur 1. Par sa forme, ce mode de sélection correspond à la sélection par familles. - Deuxième cas : le degré de sélection est infiniment petit et par fratrie on ne considère qu'un seul porteur de caractère qui tient lieu de proband (sélection simple, « single selection »). Il paraît évident qu'on doit, pour l'interprétation des observations, faire certaines hypothèses simplificatrices en ce qui concerne les conditions de

sélection et qui permettent de faire des constatations utilisables sur la fréquence des caractères, si elles sont appliquées avec pondération et en tenant compte de la réalité.

On peut se livrer à des considérations analogues aux précédentes dans le cas plus général où l'on ne considère que les fratries avec deux sujets atteints au minimum, au lieu de fratries avec au minimum un porteur. Parmi les fratries avec au minimum un porteur de caractère, on en trouve, cela va sans dire, qui présentent exactement un porteur de caractère ; mais on peut penser que les cas qui apparaissent isolément échappent plus facilement à l'observation, ou du moins ne retiennent-ils pas l'attention autant que des fratries où un caractère se manifeste plusieurs fois. Ainsi il est possible qu'un échantillon de fratries comporte trop peu de fratries avec un sujet atteint. L'utilisation exclusive de fratries avec au minimum deux porteurs de caractère semble également indiquée lorsqu'il existe, à côté de la forme héréditaire du caractère, une forme non héréditaire apparaissant de manière sporadique (phénocopie). Dans le rétinoblastome, par exemple, on ne peut pas distinguer d'une façon certaine, cliniquement et histologiquement, la forme héréditaire du caractère de la forme sporadique. Il est nécessaire, en présence d'antécédents sans particularités, de se limiter à des fratries avec au minimum deux porteurs de caractère (ou tarés), pour s'assurer qu'il ne s'agit que de cas héréditaires.

On se limite à la sélection simple et complète, en plus on renonce à la description des procédés de comparaison et on se borne aux procédés d'estimation de la fréquence relative des tarés. On évitera de décrire ici les procédés de comparaison, qui ont été développés par WEINBERG (1912 ff), APERT (1914), JUST (1920), BERNSTEIN (1925 ff), LENZ (1929 ff) et d'autres, et qui servent en premier lieu à examiner une hypothèse héréditaire donnée (fréquence de caractère dans l'ensemble de base, par exemple égal à 25 pour cent). On s'en passera, premièrement, parce qu'on peut vérifier immédiatement une hypothèse héréditaire déterminée en faisant l'estimation de la fréquence de caractère, pour autant que celle-ci s'accompagne de l'erreur standard qui lui est propre, de sorte qu'en principe, le procédé d'estimation est en même temps un procédé de vérification ; deuxièmement, parce que ces procédés d'estimation aboutissent (si certaines conditions de sélection sont remplies) là aussi à des conclusions utiles pour des caractères dominants incomplètement pénétrants où la simple vérification d'une hypothèse héréditaire serait plus ou moins superflue.

Les procédés d'estimation dont il sera question plus bas, sont ceux de WEINBERG, RIDER, MOORE et celui du «Maximum Likelihood» d'après FISHER et les «Maximum Likelihood Scores» de FINNEY. Ces différents procédés d'estimation ne concernent que l'analyse des observations et leur emploi pour obtenir une caractéristique statistique qui réponde aussi exacte-

ment que possible aux différentes questions. Ils ne touchent pas le mode de sélection, mais ils se basent plutôt sur le même modèle statistique adapté à des conditions de sélections bien déterminées, à savoir celui des répartitions binomiales tronquées.

Dans un groupe de fratries avec porteurs d'un caractère déterminé, on trouvera, en général, des fratries de différentes grandeurs; à l'intérieur des fratries de même grandeur, le nombre des tarés variera à son tour. Puisqu'on suppose que la fréquence relative des tarés dans la population est la même pour des fratries de grandeur variable, on s'efforcera de faire une estimation de la fréquence de caractère qui tienne compte des fratries de différentes grandeurs. La solution de ce problème est obtenue de façon assez différente dans les divers procédés utilisés jusqu'ici et elle dépend des suppositions supplémentaires que l'on fait sur la répartition par grandeur des fratries, si l'on considère ces grandeurs mêmes comme variables stochastiques. Ainsi VON BEHR (1927) emploie, pour examiner la méthode de WEINBERG, entre autres une répartition a priori pour la fréquence relative des différentes grandeurs des fratries, dont la répartition réellement observée du nombre d'enfants doit être conçue comme un échantillon. BAILEY (1951) considère comme variable aléatoire le nombre de fratries observées d'une grandeur déterminée avec un nombre défini de tarés, dont un certain nombre sont des probands, et le nombre total des fratries d'une grandeur déterminée avec au minimum un sujet atteint apparaît ainsi également comme variable aléatoire qui permet de tirer des conclusions sur le chiffre total de fratries de cette grandeur dans la population des fratries avec porteurs de caractère réels et virtuels. Il ressort immédiatement du procédé employé dans les deux cas que nous venons de traiter, comment on peut, en se basant sur des fratries de toutes grandeurs, parvenir à une estimation de la fréquence de caractère. Dans le présent travail, en accord avec des auteurs plus récents tels que FISHER (1934), HALDANE (1938, 1947-49) et FINNEY (1947-49), nous ne supposons pas que la répartition de l'ensemble de base du nombre d'enfants en fratries reste la même d'échantillon en échantillon; on s'appuiera plutôt sur une répartition observée des grandeurs de fratries; et pour une grandeur de fratries déterminée, le caractère aléatoire n'apparaîtra que dans le nombre de tarés dans chacune des fratries. On prendra donc, comme ensemble de base, la répartition binomiale tronquée, adaptée aux conditions de sélection, du nombre de porteurs de caractère dans des fratries d'une certaine grandeur déterminée par une fréquence de caractère p à estimer à partir des observations. Il faut remarquer que les méthodes de MOORE et de RIDER ne se rapportent, à l'origine, qu'à l'estimation de p sur la base d'échantillons de grandeur s suivant une répartition binomiale tronquée; le cas de l'effectif d'échantillon variable, tel qu'il se présente en génétique médicale, n'y est pas traité du tout. Dans les suppositions faites quant au nombre de fratries de différentes grandeurs, la meilleure solution du problème d'estimation de la fréquence de caractère, compte tenu de toutes les grandeurs de fratries, consiste à estimer la fréquence de caractère pour chacune des grandeurs des fratries, et à utiliser comme estimation combinée une

moyenne pondérée des estimations individuelles, en prenant pour poids l'inverse de la variance des estimations. Pour fixer une fréquence de caractère commune p dans des fratries de grandeur quelconque, on n'utilisera cependant pas comme poids la variance calculée pour les valeurs des estimations individuelles, mais on calculera la variance toujours pour la même valeur et l'on fera varier cette valeur jusqu'à ce que l'estimation pondérée coïncide avec la valeur pour laquelle les poids ont été calculés. Cette valeur servira d'estimation combinée de la fréquence de caractère. Ce procédé-ci diffère considérablement de celui de WEINBERG, se rallie plutôt à la méthode des scores. C'est probablement le seul procédé adéquat pour évaluer de façon objective les estimations individuelles et pour obtenir des estimations globales à peu près uniformes pour les différents procédés d'estimation que nous venons d'énumérer.

Le présent travail a pour but de comparer, au moyen de critères statistiques, les dites méthodes d'estimation selon WEINBERG, RIDER et MOORE et celle du Maximum Likelihood. L'application de ces méthodes pourra être facilitée par l'emploi des tables ci-jointes donnant les poids pour des valeurs p de la fréquence théorique du caractère suffisamment rapprochées ; quelques exemples numériques tirés de la génétique humaine en illustrent les calculs pratiques.

II. QUELQUES NOTIONS FONDAMENTALES DE STATISTIQUE

Ainsi que le montrent les lois de l'hérédité, la transmission du patrimoine héréditaire des parents à leurs enfants est régie par des processus aléatoires qui peuvent être décrits dans certains cas par une répartition binomiale. Ceci est surtout valable pour les caractères héréditaires mendéliens simples dominants, récessifs ou intermédiaires, où l'on peut observer chez les descendants les proportions connues 3:1, 1:1, 1:2:1, à condition qu'on puisse grouper les parents en classes de même constitution héréditaire. L'observation de ces proportions est liée à la condition que, dans la première génération de descendants, on puisse inclure des fratries qui ne présentent pas de porteurs de caractère, mais qui auraient pu en présenter d'après la constitution génétique de leurs parents. Pour des caractères récessifs rares, cette condition ne peut pas, en général, être remplie parce que le patrimoine héréditaire récessif ne peut se manifester habituellement que chez les descendants de deux parents hétérozygotes. Des fratries de parents hétérozygotes sans porteur de caractère échappent d'emblée à l'observation, et il en résulte un effet de sélection, dans le sens d'un nombre trop élevé de récessifs. Pour examiner malgré tout si une proportion théorique de un sujet atteint sur

trois sujets non-atteints est valable, il faut corriger cet effet de sélection. C'est ce qu'on fera par la suite. Afin de tenir compte, en même temps du mode de sélection, on procédera d'une manière un peu plus générale qu'il ne serait absolument nécessaire pour les développements ultérieurs.

On part de fratries de grandeur s avec parents non atteints et un porteur de caractère au minimum. Soit p la fréquence théorique des sujets atteints, c'est-à-dire la fréquence avec laquelle les zygotes responsables de la manifestation du caractère chez l'individu en formation, sont formés par la fusion des gamètes des parents lors du processus aléatoire de la rencontre des patrimoines héréditaires des parents. Pour tenir compte de l'influence du mode de sélection, on désignera par k la probabilité qu'un porteur de caractère devienne proband. Puis nous calculerons la probabilité qu'une fratrie de grandeur s présente r porteurs de caractère, dont t probands, à condition qu'on ne prenne en considération que des fratries avec au moins un sujet atteint et au moins un proband. Pour un membre d'une telle fratrie, il y a trois possibilités : il est soit non-atteint avec la probabilité $q = 1 - p$, soit atteint et non proband avec la probabilité $p(1 - k) = pl$, soit atteint et proband avec la probabilité pk . La probabilité qu'un membre soit normal, ou atteint et non proband, est donc $q + p(1 - k) = 1 - kp$; la probabilité que les s membres d'une fratrie soient normaux ou non probands est $(1 - kp)^s$, et la probabilité qu'un membre au moins soit atteint et proband est dès lors $1 - (1 - kp)^s$. La probabilité conditionnelle P_{srt} qu'une fratrie de grandeur s qui est observable, c'est-à-dire susceptible d'être prise dans l'échantillon présente r sujets atteints, dont t probands, résulte de la répartition multinomiale

$$P_{srt} = \frac{1}{1 - (1 - kp)^s} \frac{s!}{t! (r - t)! (s - r)!} (pk)^t (pl)^{r - t} q^{s - r} \quad (2.1)$$

$$1 \leq t \leq r \leq s$$

De même, on peut calculer la probabilité qu'une fratrie de grandeur s contienne r sujets atteints dont exactement t probands, dans le cas où l'on se limite d'emblée à l'observation de fratries avec, au minimum, deux sujets atteints et un proband. Ce cas se présente lorsque, par suite de hasards dans la sélection des fratries avec un seul sujet atteint seulement, la garantie est minime que ces fratries avec un seul individu atteint soient représentées avec la bonne fréquence; ou bien lorsqu'on peut attribuer des cas isolés dans des fratries à des causes non héréditaires. Dans tous les cas, il faut s'attendre à une disproportion entre le nombre de fratries avec un seul sujet atteint et de celles avec plusieurs individus atteints, disproportion qui fausserait l'estimation de la fréquence du caractère. La probabilité qu'une fratrie de grandeur s présente un nombre quelconque de sujets atteints, mais aucun proband, vaut ainsi que nous venons de la calculer $(1 - kp)^s$; la probabilité qu'une fratrie présente exactement un sujet atteint et proband et $(s - 1)$ membres normaux, est $skpq^{s-1}$; donc, la probabilité complémentaire $1 - (1 - kp)^s - skpq^{s-1}$ est la probabilité qu'une fratrie contienne au moins deux sujets atteints et au

moins un proband. La probabilité P_{srt} qu'une fratrie pouvant être observée dans les conditions mentionnées, présente r sujets atteints dont t probands, est dans ce cas alors

$$P_{srt} = \frac{1}{1 - (1 - kp)^s - skpq^{s-1}} \frac{s!}{t! (r-t)! (s-r)!} (pk)^t (pl)^{r-t} q^{s-r} \quad (2.2)$$

$$1 \leq t \leq r, \quad 2 \leq r \leq s$$

Les deux probabilités calculées sont importantes pour l'analyse de fratries suivant la méthode par probands à proprement parler, lorsqu'on dispose de fratries avec plus de 1 mais moins de s probands. En général, les indications concernant le nombre de probands, sélectionnés indépendamment les uns des autres dans des fratries, sont peu sûres, et quoiqu'on puisse déduire simultanément p et k des observations, les possibilités d'appliquer des procédés d'estimation correspondants sont plutôt rares. Les deux cas limites $k \rightarrow 0$, $l \rightarrow 1$ et $k \rightarrow 1$, $l \rightarrow 0$ sont d'un intérêt plus pratique.

Un simple passage à la limite montre que les probabilités indiquées ci-dessus ne convergent vers une limite non identiquement nulle que si l'on pose $t = 1$ pour $k \rightarrow 0$, et $t = r$ pour $k \rightarrow 1$. Le cas extrême d'une probabilité de sélection très petite correspond donc au cas de la sélection simple (single selection) avec un seul proband, et le cas limite de la probabilité de sélection 1 au cas de la sélection complète (complete selection), tous les sujets atteints étant considérés comme probands. La probabilité P_{srt} prend, dans ces conditions, la forme suivante :

1) Nombre de sujets atteints dans des fratries observables égal à 1 au moins ; $\rho = 1$:

$$k = 0 : P_{sr} = \binom{s-1}{r-1} p^{r-1} q^{s-r} \quad 1 \leq r \leq s \quad (2.3)$$

$$k = 1 : P_{sr} = \frac{1}{1 - q^s} \binom{s}{r} p^r q^{s-r} \quad (2.4)$$

2) Nombre de sujets atteints dans des fratries observables égal à 2 ; $\rho = 2$:

$$k = 0 : P_{sr} = \frac{1}{1 - q^{s-1}} \binom{s-1}{r-1} p^{r-1} q^{s-r} \quad 2 \leq r \leq s \quad (2.5)$$

$$k = 1 : P_{sr} = \frac{1}{1 - q^s - spq^{s-1}} \binom{s}{r} p^r q^{s-r} \quad (2.6)$$

Si l'on multiplie ces probabilités par le nombre observé n_s de fratries de grandeur s , on obtiendra le nombre théorique de fratries de grandeur s , avec r sujets atteints, auquel il faut s'attendre dans le cas du procédé de sélection en question.

Pour $k = 1$, c'est-à-dire en supposant que tous les sujets atteints soient probands, on a alors, pour probabilité qu'une fratrie présente r sujets atteints, une répartition binomiale tronquée une fois ou deux fois suivant que $\rho = 1$ ou $\rho = 2$. Pour $k = 0$, on a, pour $\rho = 1$, une répartition binomiale non-tronquée pour probabilité d'apparition du caractère parmi les frères et sœurs du proband. La répartition n'est pas tronquée parce que, si l'on fait abstraction du proband dans les fratries, le procédé de sélection fournit, sous forme de fratries avec un seul sujet atteint, les fratries «vides» nécessaires à l'estimation de la fréquence de caractère, même s'il s'agit de fratries vides de grandeur $s-1$. De même, pour $k' = 0$, on obtient, pour des fratries observables avec au minimum deux atteints ($\rho = 2$), une répartition binomiale tronquée une fois seulement au lieu d'une répartition tronquée deux fois lorsque $k = 1$; on considère de nouveau les fratries réduites de l'unique proband, et puisqu'on n'observe que les fratries avec, au minimum, deux sujets atteints, la répartition est alors tronquée une fois. La grandeur des fratries devant être prises en considération, de même que le nombre des sujets atteints dont il faut tenir compte diminuent comme précédemment de s à $s-1$ et de r à $r-1$ respectivement.

Pour la sélection complète ($k = 1$), on considérera la fratrie complète, pour la sélection simple ($k = 0$), la fratrie sans le proband unique. Pour la sélection complète, les répartitions de probabilité pour des fratries de grandeur s avec r sujets atteints consistent en les répartitions binomiales ordinaires tronquées une fois ou deux fois; le passage à la sélection simple s'effectue en substituant à la répartition binomiale tronquée une fois ou deux fois la répartition non tronquée ou tronquée une fois respectivement, et l'on remplacera en même temps s par $s-1$, et r par $r-1$. Ces constatations permettent de définir, pour des procédés d'estimation développés, comme ceux de RIDER et MOORE, à partir de l'hypothèse implicite de $k = 1$, les procédés d'estimation correspondants à $k = 0$; ceci est en particulier valable pour l'estimation de la fréquence de caractère de fratries avec au moins deux sujets atteints, si l'on suppose que $k = 0$; dans ce cas, on utilisera le procédé d'estimation relatif à des fratries avec un sujet atteint au moins, et pour $k = 1$ en substituant, bien entendu, $s-1$ et $r-1$ à s et r respectivement.

Espérance mathématique du nombre de sujets atteints.

$$\begin{aligned}
 \rho = 1, k = 1. E_s(r) &= \frac{1}{1 \cdot q^s} \sum_{r=1}^s r \binom{s}{r} p^r q^{s-r} \\
 &= \frac{1}{1 \cdot q^s} s p \sum_{r=1}^s \binom{s-1}{r-1} p^{r-1} q^{s-r} \\
 &= \frac{sp}{1 \cdot q^s}
 \end{aligned}
 \tag{2.7}$$

$$\begin{aligned}
 k = 0. E_s(r) &= \sum_{r=1}^s r \binom{s-1}{r-1} p^{r-1} q^{s-r} \\
 &= (s-1)p \sum_{r=1}^s \binom{s-2}{r-2} p^{r-2} q^{s-r} + \sum_{r=1}^s \binom{s-1}{r-1} p^{r-1} q^{s-r} \\
 &= 1 + (s-1)p
 \end{aligned} \tag{2.8}$$

$(s-1)p$ est l'espérance mathématique du nombre d'individus atteints parmi les frères et sœurs du proband; 1 est mis pour le proband lui-même. Si l'on écrit $E_s(r-1) = (s-1)p$, la règle énoncée ci-dessus est valable pour le passage de $k=1$ à $k=0$.

$$\begin{aligned}
 \rho = 2, k = 1 : E_s(r) &= \frac{1}{1 - q^s - spq^{s-1}} \sum_{r=2}^s r \binom{s}{r} p^r q^{s-r} \\
 E_s(r) &= \frac{sp}{1 - q^s - spq^{s-1}} \sum_{r=2}^s \binom{s-1}{r-1} p^{r-1} q^{s-r} \\
 &= \frac{sp(1 - q^{s-1})}{1 - q^s - spq^{s-1}}
 \end{aligned} \tag{2.9}$$

$$\begin{aligned}
 k = 0 : E_s(r) &= 1 + E_s(r-1) \\
 &= 1 + \frac{1}{1 - q^{s-1}} \sum_{r=2}^s (r-1) \binom{s-1}{r-1} p^{s-1} q^{s-r} \\
 &= 1 + \frac{(s-1)p}{1 - q^{s-1}} \sum_{r=2}^s \binom{s-2}{r-2} p^{s-2} q^{s-r} \\
 &= 1 + \frac{(s-1)p}{1 - q^{s-1}}
 \end{aligned} \tag{2.10}$$

Pour un s déterminé, le nombre r de sujets atteints dans une fratrie est une variable aléatoire, dont la répartition ne dépend que de la fréquence théorique p et de certaines hypothèses particulières sur la sélection de la fratrie. Si l'on a un échantillon comptant n_s fratries de grandeur s , cet effectif n_s ne devra donc pas, suivant notre hypothèse, être considéré comme une variable aléatoire, mais comme une grandeur fixe, et l'espérance mathématique $E(r_s)$ du nombre total r_s de sujets atteints dans les n_s fratries sera égal à la somme des espérances mathématiques pour les fratries prises une à une. Ce qui donne en résumé :

$$\rho = 1, k = 1 : E(r_s) = n_s \frac{sp}{1 - q^s} \tag{2.11}$$

$$\rho = 1, k = 0 : E(r_s) = n_s [1 + (s-1)p] \tag{2.12}$$

$$\rho = 2, k = 1: E(r_s) = n_s \frac{sp(1-q^{s-1})}{1-q^s - spq^{s-1}} \quad (2.13)$$

$$\rho = 2, k = 0: E(r_s) = n_s \left[1 + \frac{(s-1)p}{1-q^{s-1}} \right] \quad (2.14)$$

Variance du nombre r de sujets atteints.

On peut calculer la variance $V_s(r)$ du nombre r de sujets atteints dans une fratrie de grandeur s par l'équation

$$V_s(r) = E_s(r^2) - [E_s(r)]^2 \quad (2.15)$$

Pour $\rho = 1, k = 1$, on a

$$\begin{aligned} E_s(r^2) &= E_s[r(r-1)] + E_s(r) \\ &= \frac{1}{1-q^s} \sum_{r=1}^s r(r-1) \binom{s}{r} p^r q^{s-r} + E_s(r) \\ &= \frac{1}{1-q^s} s(s-1)p^2 \sum_{r=1}^s \binom{s-2}{r-2} p^{r-2} q^{s-r} + E_s(r) \\ &= \frac{s(s-1)p^2}{1-q^s} + \frac{sp}{1-q^s} \\ V_s(r) &= \frac{s(s-1)p^2}{1-q^s} + \frac{sp}{1-q^s} - \frac{s^2 p^2}{(1-q^s)^2} \\ &= \frac{sp}{(1-q^s)^2} [(sp+q)(1-q^s) - sp] \\ &= \frac{sp}{(1-q^s)^2} [sp - spq^s + q - q^s + 1 - sp] \\ &= \frac{spq}{(1-q^s)^2} (1-q^s - spq^{s-1}) \end{aligned} \quad (2.16)$$

Pour $\rho = 1, k = 0$, on a de même

$$E_s(r^2) = E_s[r(r-1)(r-2)] + 3E_s(r-1) + 1$$

$$E_S(r^2) = 1 + \sum_{r=1}^s (r-1)(r-2) \binom{s-1}{r-1} p^{r-1} q^{s-r} + 3 E_S(r-1)$$

$$= 1 + (s-1)(s-2)p^2 + 3(s-1)p$$

$$V_S(r) = E_S(r^2) - [E_S(r)]^2$$

$$= 1 + (s-1)(s-2)p^2 + 3(s-1)p - [1 + (s-1)p]^2$$

$$= [(s-1)(s-2) - (s-1)^2] p^2 + (s-1)p$$

$$= (s-1)pq \quad (2.17)$$

De façon analogue, on peut calculer $V_S(r)$ lorsque $\rho = 2$, $k = 1$ ou $k = 0$. On obtient :

$$\rho = 2, k = 1 : V_S(r) = spq \frac{(1-q^{s-1})^2 - (s-1)^2 p^2 q^{s-2}}{(1-q^s - spq^{s-1})^2} \quad (2.18)$$

$$\rho = 2, k = 0 : V_S(r) = (s-1)pq \frac{1-q^{s-1} - (s-1)pq^{s-2}}{(1-q^{s-1})^2} \quad (2.19)$$

Cette dernière formule correspond, comme il était à prévoir, à celle pour laquelle $\rho = 1$, $k = 1$, lorsqu'on y remplace s par $s-1$.

Comme pour l'espérance mathématique, on peut calculer la variance $V(r_S)$ du nombre total de sujets atteints dans les n_S fratries en multipliant par n_S la variance $V_S(r)$ pour une seule fratrie, puisqu'on suppose que le nombre n_S de fratries de grandeur s est donné et non soumis au hasard. C'est-à-dire qu'on additionne simplement les variances de chacune des fratries, ce qui est permis parce que les fratries isolées doivent être considérées comme indépendantes les unes des autres conformément aux hypothèses sur le modèle décrivant la sélection des fratries. En résumé, on a :

$$\rho = 1, k = 1 : V(r_S) = n_S spq \frac{1-q^s - spq^{s-1}}{(1-q^s)^2} \quad (2.20)$$

$$\rho = 1, k = 0 : V(r_S) = n_S (s-1)pq \quad (2.21)$$

$$\rho = 2, k = 1 : V(r_S) = n_S spq \frac{(1-q^{s-1})^2 - (s-1)^2 p^2 q^{s-1}}{(1-q^s - spq^{s-1})^2} \quad (2.22)$$

$$\rho = 2, k = 0 : V(r_S) = n_S (s-1)pq \frac{1-q^{s-1} - (s-1)pq^{s-2}}{(1-q^{s-1})^2} \quad (2.23)$$

Espérances mathématiques et variances dans les répartitions tronquées et non tronquées.

Il est intéressant de comparer les variances et les espérances mathématiques, calculées plus haut pour des répartitions binomiales tronquées, à celles qu'on observe dans les répartitions non tronquées, qui seraient valables si l'on n'avait pas à tenir compte des effets de sélection. Pour la répartition binomiale non tronquée, on obtiendrait comme espérance mathématique du nombre r des sujets atteints dans n_s fratries de grandeur s et pour une fréquence théorique p , la valeur

$$E(r_s) = n_s \cdot sp,$$

ce qu'on peut aussi écrire

$$E\left(\frac{r_s}{n_s \cdot s}\right) = p,$$

puisque $n_s \cdot s$ représente le nombre total des membres de ces fratries. Le quotient $\hat{p}_s = r_s / (n_s \cdot s)$ du nombre total de sujets atteints dans les n_s fratries, compte tenu de tous leurs membres, serait donc une estimation non biaisée de la fréquence de caractère théorique p , à laquelle on pourrait associer la variance

$$V(\hat{p}_s) = V\left(\frac{r_s}{n_s \cdot s}\right) = \frac{pq}{s \cdot n_s}.$$

Si l'on doit tenir compte des conditions de sélection et, par conséquent, utiliser des répartitions binomiales tronquées, on formera, pour $k = 1$, le quotient $r_s / (n_s \cdot s)$ selon (2.11) et (2.13), et, pour $k = 0$, le quotient $(r_s - n_s) / n_s (s - 1)$ selon (2.12) et (2.14). Dans le cas où $\rho = 1$, $k = 0$ seulement, on a une estimation non biaisée de la fréquence de caractère p , à la condition cependant qu'avant la formation du quotient, les n_s probands soient soustraits du nombre de sujets atteints ainsi que du nombre total d'enfants. La variance correspondante est

$$V(\hat{p}_s) = pq / n_s (s - 1).$$

Dans les trois autres cas, $\rho = 1$, $k = 1$; $\rho = 2$, $k = 0$ ou 1 , on n'a pas p , mais p multiplié par un facteur dont on peut démontrer qu'il est plus grand que 1. C'était à prévoir, puisqu'on obtient un nombre trop élevé de récessifs ou de sujets atteints lorsqu'on ne considère, dans certaines conditions de sélection déterminées, que des fratries avec un ou deux sujets atteints au moins. Si, abstraction faite du cas de la sélection simple où $k = 0$, $\rho = 1$, l'espérance mathématique $r_s / (n_s \cdot s)$ ou $(r_s - n_s) / n_s (s - 1)$ est trop élevée par rapport à celle d'un matériel non sélectionné, pour un matériel sélectionné, au contraire, la variance des quotients $r_s / (n_s \cdot s)$ ou $(r_s - n_s) / n_s (s - 1)$ est plus petite que celle de la répartition binomiale non tronquée. La discussion ci-dessus montre que, à l'exception du cas où $k = 0$, $\rho = 1$, la fonction de l'échantillon r_s / n_s ou $(r_s - n_s) / n_s (s - 1)$ ne convient pas directement à l'estimation de la fréquence de caractère p .

III. ESTIMATION DE LA FRÉQUENCE DE CARACTÈRE p PAR IDENTIFICATION DES OBSERVATIONS ET DE L'ESPÉRANCE MATHÉMATIQUE.

MÉTHODE DU MAXIMUM LIKELIHOOD.

MÉTHODE DES SCORES.

Dans la littérature génétique, on cite différents procédés couramment utilisés pour la vérification d'une hypothèse héréditaire sur les données d'un matériel d'observation; il faut notamment mentionner la méthode directe ou a priori, ou le procédé de BERNSTEIN, le procédé de MACKLIN et le procédé du nombre total théorique d'enfants, de LENZ et HOGBEN. La méthode la plus indiquée consiste à calculer une estimation de la fréquence de caractère p , et à examiner si cette estimation s'écarte significativement ou non de la valeur hypothétique.

Tout d'abord, on ne considérera de nouveau que des fratries d'une grandeur définie s . Soit n_s leur nombre. En supposant qu'on n'observe que des fratries avec, au minimum, un sujet atteint ($\rho = 1$), ces fratries représentent un échantillon d'effectif n_s ; et pour chaque fratrie, on aura un nombre r de sujets atteints situé entre 1 et s . Le nombre r des sujets atteints doit être considéré comme variable aléatoire dont la répartition de fréquence est fixée par les relations (2.3) à (2.6).

Par fonction d'estimation, on entend une fonction appropriée h des valeurs des variables aléatoires dans l'échantillon; elle peut être considérée comme une approximation du paramètre à estimer. Pratiquement, il s'agit donc d'une fonction du nombre de sujets atteints dans les n_s fratries, fonction qui s'approche suffisamment de la valeur p de l'ensemble de base dans lequel on suppose avoir prélevé les échantillons.

Méthode d'identification de l'observation et de l'espérance mathématique.

A la fin du chapitre précédent, on a constaté que la fréquence relative $r_s / (n_s \cdot s)$ des sujets atteints dans les fratries de grandeur s ne représente pas, en général, une estimation appropriée de la fréquence de caractère théorique p , si l'on doit tenir compte des effets de sélection. Mais on peut obtenir immédiatement une estimation de p en égalant les nombres observés aux nombres théoriques des sujets atteints, et en considérant la relation résultante comme équation déterminant l'estimation $\hat{p}_s$.

En désignant par r_s le nombre des sujets atteints qu'on a observés dans les n_s fratries, et en posant

$$\bar{r}_s = r_s / n_s \quad (3.1)$$

on aura immédiatement de (2.11) à (2.14) les équations

$$\rho = 1, k = 0: \bar{r}_s - 1 = (s-1) p \quad (3.2)$$

$$k = 1: \bar{r}_s = \frac{sp}{1-q^s} \quad (3.3)$$

$$\rho = 2, k = 0: \bar{r}_s - 1 = \frac{(s-1)p}{1-q^{s-1}} \quad (3.4)$$

$$k = 1: \bar{r}_s = \frac{sp(1-q^{s-1})}{1-q^s - spq^{s-1}} \quad (3.5)$$

En ce qui concerne les solutions $\hat{p}_s$, il s'agit manifestement de fonctions h' qui ne dépendent que des valeurs r des variables aléatoires dans l'échantillon.

On peut sans autre résoudre l'équation dans le cas où $\rho = 1, k = 0$; on obtient :

$$\hat{p}_s = \frac{\bar{r}_s - 1}{s-1} = \frac{r_s - n_s}{n_s (s-1)} \quad (3.6)$$

Cette estimation correspond à la méthode d'estimation de la fréquence de caractère indiquée par JUST (1920). On s'aperçoit que, pour $k = 0$, on soustrait 1 de $\bar{r}_s$ et de s , conformément au fait qu'on n'a pas tenu compte du proband.

Afin qu'une estimation soit utilisable, on exige d'elle certaines propriétés. Il est d'abord important que l'estimation soit correcte (« consistente »), c'est-à-dire que pour un effectif d'échantillon n_s croissant, la valeur estimée du paramètre en question converge stochastiquement vers la valeur de ce paramètre dans l'ensemble de base. Pour examiner la validité de ces propriétés, il est nécessaire de regarder d'un peu plus près les solutions des équations ci-dessus.

Pour ce qui en est tout d'abord du cas où $\rho = 1, k = 0$, la variable aléatoire individuelle r a une espérance mathématique de $1 + (s-1)p$, suivant (2.8); donc $(r-1)/(s-1)$ a une espérance mathématique de p . En tant que moyenne de grandeurs $(r-1)/(s-1)$, l'estimation $\hat{p}_s$ tend, avec un effectif d'échantillon n_s croissant, vers p en vertu de la loi des grands nombres; c'est-à-dire que $\hat{p}_s$ est même non biaisé, autrement dit, on a pour tout n_s

$$E(\hat{p}_s) = \frac{E(r_s) - n_s}{n_s (s-1)} = p$$

En guise d'exemple représentatif des autres cas, nous analyserons le cas où $\rho = 1, k = 1$. $\hat{p}_s$ est alors la solution de l'équation

$$\frac{r_s}{n_s \cdot s} = \frac{\hat{p}_s}{1 - (1 - \hat{p}_s)^s}.$$

En supposant qu'il s'agisse d'un ensemble de base avec le paramètre p , r_s / n_s a pour espérance mathématique $p / (1 - q^s)$;

en soustrayant $p / (1 - q^s)$ des deux membres de l'équation, on a

$$\frac{r_s}{n_s} - \frac{p}{1 - (1-p)^s} = \frac{\hat{p}_s}{1 - (1-\hat{p}_s)^s} - \frac{p}{1 - (1-p)^s}.$$

Pour l'écart entre $r_s / n_s = \bar{r}_s / s$ et son espérance mathématique, nous poserons $\Delta (r_s / s)$; l'écart $\hat{p}_s - p$ sera désigné par Δp ; de plus, pour simplifier $p / (1 - q^s) = f(p)$.

Le développement en série de puissance de Δp donne :

$$\Delta \left(\frac{\bar{r}_s}{s} \right) = f'(p) \Delta p + \frac{1}{2!} f''(p) (\Delta p)^2 + \dots$$

Ce développement en série peut être inversé, de façon à ce qu'on obtienne, pour Δp :

$$\Delta p = \frac{1}{f'(p)} \Delta \left(\frac{\bar{r}_s}{s} \right) - \frac{f''}{2!} \frac{1}{(f')^3} \left[\Delta \left(\frac{\bar{r}_s}{s} \right) \right]^2 + \dots (3.7)$$

Si p est le paramètre de l'ensemble de base dans lequel l'échantillon de fratries a été prélevé, pour n_s croissant, $\bar{r}_s / s$ s'approchera stochastiquement toujours davantage de son espérance mathématique $p / (1 - q^s)$, d'après la loi des grands nombres; en d'autres termes, $\Delta (\bar{r}_s / s)$ converge stochastiquement vers zéro; ainsi Δp tend stochastiquement vers zéro également, c'est-à-dire $\hat{p}_s$ s'approche stochastiquement de p , de sorte que l'estimation $\hat{p}_s$ est correcte.

Mais alors l'estimation $\hat{p}_s$ est biaisée, car $E [\Delta (\bar{r}_s / s)] = 0$, et dès lors

$$E(\Delta p) = E(\hat{p}_s) - p = - \frac{f''}{2!} \frac{1}{(f')^3} E [\Delta (\bar{r}_s / s)]^2 + \dots$$

Cependant,

$$E \left[\Delta \left(\frac{\bar{r}_s}{s} \right) \right] = V \left(\frac{\bar{r}_s}{s} \right) = \frac{1}{n_s^2} V(r_s) = \frac{V_s(r)}{n_s s^2} > 0 \quad (3.8)$$

et également $f' > 0$, il s'ensuit que

$$E(\hat{p}_s) \neq p,$$

c'est-à-dire que, si l'on se limite aux termes d'ordre n_s^{-1} , l'estimation $\hat{p}_s$ est entachée d'une erreur systématique, négative ou positive, suivant le signe de f'' . L'estimation $\hat{p}_s$ est donc biaisée. Dans le cas considéré où $\rho = 1$, $k = 1$,

l'erreur systématique sera

$$E(\Delta p) = -\frac{1}{n_s} \frac{pq^{s-1}(1-q^s)}{2(1-q^s - spq^{s-1})^2} [(s+1)p - 2 + \{(s-1)p + 2\}q^s] \quad (3.9)$$

On aboutit à des expressions semblables ou plus compliquées pour les deux cas restants $\rho = 2, k = 0$ ou $k = 1$.

Si l'on élève cette relation (3.7) au carré et qu'on passe à l'espérance mathématique, en négligeant les puissances supérieures à 2 en $\Delta(\hat{r}_s/s)$, on obtient dans les cas de $\rho = 1, k = 1$,

$$E(\Delta p)^2 = V(\hat{p}_s) = \frac{1}{(f')^2} V\left(\frac{\hat{r}_s}{s}\right) = \frac{1}{(f')^2} \frac{V_s(r)}{n_s s^2},$$

et comme

$$V_s(r) = \frac{spq}{(1-q^s)} (1-q^s - spq^{s-1}) \text{ et } f'(p) = \frac{1-q^s - spq^{s-1}}{(1-q^s)^2}$$

finalement

$$V(\hat{p}_s) = -\frac{1}{n_s} \frac{pq}{s} \frac{(1-q^s)^2}{(1-q^s - spq^{s-1})}.$$

Dans les autres cas, on obtient des formules analogues qui peuvent être groupées de la manière suivante :

$$\rho = 1, k = 0: V(\hat{p}_s) = \frac{pq}{n_s (s-1)} \quad (3.10)$$

$$\rho = 1, k = 1: V(\hat{p}_s) = \frac{pq}{s n_s} \frac{(1-q^s)^2}{(1-q^s - spq^{s-1})} \quad (3.11)$$

$$\rho = 2, k = 0: V(\hat{p}_s) = \frac{pq}{(s-1)n_s} \frac{(1-q^{s-1})^2}{(1-q^s - (s-1)pq^{s-2})} \quad (3.12)$$

$$\rho = 2, k = 1: V(\hat{p}_s) = \frac{pq}{s n_s} \frac{(1-q^s - spq^{s-1})^2}{(1-q^{s-1})^2 - (s-1)^2 p^2 q^{s-2}} \quad (3.13)$$

Par principe, on remplacera p par la valeur estimée. Pour $\rho = 1, k = 0$, on obtient la même variance que pour une répartition binomiale non tronquée de l'effectif d'échantillon $n_s (s-1)$. Dans les autres cas, la variance est plus élevée en comparaison avec celle de la répartition non tronquée.

Le procédé du Maximum Likelihood.

Un procédé important pour l'estimation des paramètres d'un ensemble de base est celui du Maximum Likelihood de R.A. FISHER (vraisemblance maximum). Si, tout d'abord, on se limite à nouveau aux fratries de grandeur s , et qu'on suppose que n_s familles de cette grandeur ont été observées, on obtient une certaine répartition observée n_{sr} des n_s fratries en classes de diverses valeurs r , répartition donnée par l'expression

$$P_s = \frac{(n_s)!}{\prod (n_{sr})!} \prod_r (P_{sr})^{n_{sr}}, \quad (3.14)$$

dans laquelle on calcule P_{sr} par l'une des formules (2.3) - (2.6), le produit $\prod$ s'étendant au domaine correspondant des valeurs r . P_s représente une fonction définie du paramètre p à estimer; son logarithme $\ln P_s$, considéré comme fonction de p , est appelé « fonction de vraisemblance » (Likelihood Function):

$$L_s = \ln P_s = \text{const.} + \sum_{(r)} n_{sr} P_{sr} \quad (3.15)$$

En substituant à P_{sr} les expressions (2.3) - (2.6), on trouve

$$\rho = 1, k = 0 : L_s = (r_s - n_s) \ln p + (sn_s - r_s) \ln q \quad (3.16)$$

$$\rho = 1, k = 1 : L_s = -n_s \ln (1 - q^s) + r_s \ln p + (sn_s - r_s) \ln q \quad (3.17)$$

$$\rho = 2, k = 0 : L_s = -n_s \ln (1 - q^{s-1}) + (r_s - n_s) \ln p + (sn_s - r_s) \ln q \quad (3.18)$$

$$\rho = 2, k = 1 : L_s = -n_s \ln (1 - q^s - spq^{s-1}) + r_s \ln p + (sn_s - r_s) \ln q \quad (3.19)$$

L'estimation par la méthode du Maximum Likelihood consiste alors à choisir pour estimation la valeur de p qui rend maximum la fonction P_s ou $\ln P_s$, ce qui revient au même. On dérivera alors $L = \ln P$ par rapport à p , et on égalera cette dérivée à zéro. On aura

$$\begin{aligned} \rho = 1, k = 0 : \frac{\partial L_s}{\partial p} &= \frac{r_s - n_s}{p} - \frac{sn_s - r_s}{1 - p} \\ &= \frac{r_s - n_s}{pq} - \frac{n_s (s - 1)}{q} = 0 \end{aligned} \quad (3.20)$$

$$\begin{aligned} \rho = 1, k = 1 : \frac{\partial L_s}{\partial p} &= \frac{-sn_s q^{s-1}}{1 - q^s} + \frac{r_s}{p} - \frac{sn_s - r_s}{1 - p} \\ &= \frac{r_s}{pq} - \frac{sn_s}{(1 - q^s) q} = 0 \end{aligned} \quad (3.21)$$

$$\begin{aligned} \rho = 2, k = 0 : \frac{\partial L_s}{\partial p} &= \frac{-(s-1) n_s q^{s-2}}{1-q^{s-1}} + \frac{r_s - n_s}{p} - \frac{sn_s - r_s}{1-p} \\ &= \frac{r_s - n_s}{pq} - \frac{n_s (s-1)}{(1-q^{s-1}) q} = 0 \end{aligned} \quad (3.22)$$

$$\begin{aligned} \rho = 2, k = 1 : \frac{\partial L_s}{\partial p} &= \frac{-sn_s (s-1) pq^{s-2}}{1-q^s - spq^{s-1}} + \frac{r_s}{p} - \frac{sn_s - r_s}{1-p} \\ &= \frac{r_s}{pq} - \frac{sn_s (1-q^{s-1})}{(1-q^s - spq^{s-1}) q} = 0 \end{aligned} \quad (3.23)$$

On s'aperçoit que ces expressions concordent avec les équations (3.2) - (3.5) au facteur $1/q$ près. C'est pourquoi les solutions sont les mêmes également. Elles ont la propriété d'être correctes (« consistent ») et, exception faite du cas où $\rho = 1, k = 0$, d'être biaisées.

En vertu de la théorie du Maximum Likelihood, les estimations $\hat{p}_s$ sont non seulement correctes, mais encore, lorsque n_s est grand, réparties normalement autour d'une moyenne p (vraie valeur du paramètre), avec comme variance

$$V(\hat{p}_s) = \frac{1}{V\left(\frac{\partial L_s}{\partial p}\right)} \quad (3.24)$$

où, en plus

$$V\left(\frac{\partial L_s}{\partial p}\right) = -E\left(\frac{\partial^2 L_s}{\partial p^2}\right) \quad (3.25)$$

Aucune autre variance n'est inférieure à cette dernière, et c'est pourquoi les estimations par la méthode du Maximum Likelihood méritent le qualificatif « efficace » (« efficient »). En utilisant les relations (3.20) - (3.23), on calcule tout d'abord $-\partial^2 L_s / \partial p^2$, où il faut remplacer p par la valeur $\hat{p}_s$; de là, en substituant à r_s son espérance mathématique, on tire :

$$\begin{aligned} \rho = 1, k = 0 : -\frac{\partial^2 L_s}{\partial p^2} &= \frac{1}{q} \left[\frac{r_s - n_s}{p^2} \right], \\ E\left[-\frac{\partial^2 L_s}{\partial p^2}\right] &= \frac{n_s (s-1)}{pq} \end{aligned} \quad (3.26)$$

$$\begin{aligned} \rho = 1, k = 1 : -\frac{\partial^2 L_s}{\partial p^2} &= \frac{1}{q} \left[\frac{r_s}{p^2} - \frac{n_s s^2 q^{s-1}}{(1-q^s)^2} \right], \\ E\left[-\frac{\partial^2 L_s}{\partial p^2}\right] &= \frac{n_s s}{pq} \frac{1-q^s - spq^{s-1}}{(1-q^s)^2} \end{aligned} \quad (3.27)$$

$$\rho = 2, k = 0 : - \frac{\partial^2 L_s}{\partial p^2} = \frac{1}{q} \left[\frac{r_s - n_s}{p^2} - \frac{n_s (s-1)^2 q^{s-1}}{(1-q^s)^2} \right],$$

$$E \left[- \frac{\partial^2 L_s}{\partial p^2} \right] = \frac{n_s (s-1)}{pq} \frac{1 - q^{s-1} - (s-1) p q^{s-2}}{(1 - q^{s-1})^2} \quad (3.28)$$

$$\rho = 2, k = 1 : - \frac{\partial^2 L_s}{\partial p^2} = \frac{1}{q} \left[\frac{r_s}{p^2} + \frac{n_s s (s-1) q^{s-2} (1 - sp - q^s)}{(1 - q^s - spq^{s-1})^2} \right],$$

$$E \left[- \frac{\partial^2 L_s}{\partial p^2} \right] = \frac{n_s s}{pq} \frac{(1 - q^{s-1})^2 - (s-1)^2 p^2 q^{s-2}}{(1 - q^s - spq^{s-1})^2} \quad (3.29)$$

Ces formules, déduites sous une forme un peu différente déjà par HALDANE (1947/49), conduisent aux mêmes variances que le procédé d'identification de l'observation et de l'espérance mathématique.

Pour la résolution de l'équation du Likelihood, HALDANE propose d'employer des équations mises sous la forme des membres droits de (3.20) - (3.23). Pour $\rho = 1, k = 1$ par exemple, on posera

$$\frac{r_s}{p} = \frac{sn_s}{1 - q^s}$$

Si l'on substitue, à droite, une approximation p_0 , on obtiendra à gauche une approximation p_1 ; substituée à droite, l'approximation p_1 donne à gauche une approximation p_2 , etc. On peut montrer que la suite $p_0, p_1, p_2, \dots$ converge vers la solution $\hat{p}_s$ de l'équation.

L'efficacité asymptotique fait que, pour n_s très grand, la variance du paramètre estimé à partir du Maximum Likelihood atteint réellement la variance minimum que détermine l'inéquation de CRAMER-RAO; à part cela, l'estimation par la méthode du Maximum Likelihood jouit également de la propriété d'être exhaustive («suffisant»). Cela se traduit par le fait que $\hat{p}_s$ ne dépend explicitement que du nombre total r_s de sujets atteints dans des fratries de grandeur s , et non pas du nombre r d'individus atteints dans chaque fratrie r_s , et par conséquent $n\hat{p}_s$, contient toute l'information que l'échantillon peut fournir sur le paramètre p . De toutes les méthodes d'estimation qui seront discutées dans ce travail, la méthode du Maximum Likelihood est la seule à conduire à des estimations qui ont les deux propriétés d'être exhaustives d'une part, et efficaces asymptotiquement d'autre part.

D-scores.

Ainsi qu'on le déduit des formules (3.20) - (3.23), les observations, c'est-

à-dire le nombre de sujets atteints dans les différentes fratries, entrent dans l'équation du Likelihood sous une forme particulièrement simple, les contributions des différentes fratries étant simplement additionnées; de même, l'inverse de la variance de $\hat{p}_s$ est n_s fois l'inverse de la variance pour une seule fratrie. Les différentes contributions et l'inverse de la variance d'une seule fratrie sont du type

$$\rho = 1, k = 0 : d_s = \frac{r-1}{pq} - \frac{s-1}{q} = \frac{(r-1) - E_s(r-1)}{pq}$$

$$w_s = \frac{s-1}{pq} \quad (3.30)$$

$$\rho = 1, k = 1 : d_s = \frac{r}{pq} = \frac{s}{(1-q^s)q} = \frac{r - E_s(r)}{pq}$$

$$w_s = \frac{s}{pq} \frac{1 - q^s - spq^{s-1}}{(1-q^s)^2} \quad (3.31)$$

$$\rho = 2, k = 0 : d_s = \frac{r-1}{pq} - \frac{s-1}{(1-q^{s-1})q} = \frac{(r-1) - E_s(r-1)}{pq}$$

$$w_s = \frac{s-1}{pq} - \frac{1 - q^{s-1} - (s-1)pq^{s-2}}{(1-q^{s-1})^2} \quad (3.32)$$

$$\rho = 2, k = 1 : d_s = \frac{r}{pq} - \frac{s(1-q^{s-1})}{(1-q^s - spq^{s-1})q} = \frac{r - E_s(r)}{pq}$$

$$w_s = \frac{s}{pq} \frac{(1-q^{s-1})^2 - (s-1)^2 p^2 q^{s-2}}{(1-q^s - spq^{s-1})^2} \quad (3.33)$$

On désigne les d_s comme « efficient scores » au sens de RAO (1952), et les w_s comme poids correspondants pour une fratrie de grandeur s avec r sujets atteints. Pour trouver l'équation du Likelihood pour les fratries de grandeur s et le poids correspondant, il faut simplement additionner les scores et les poids pour les différentes fratries :

Total des scores :

$$D_s = \sum \frac{n_s}{1} \frac{\partial L_s}{\partial p} = \frac{r_s - n_s \delta(k)}{pq} - n_s \frac{E_s(r) - \delta(k)}{pq} \quad (3.34)$$

où $\delta(k) = \begin{cases} 1 & \text{pour } k = 0 \\ 0 & \text{pour } k = 1 \end{cases}$

$$\text{Poids total : } W_s = \sum_{1}^{n_s} w_s = n_s w_s \quad (3.35)$$

La somme s'étendra donc aux n_s fratries de grandeur s .

Si on laisse de côté momentanément la possibilité de combiner de cette façon des fratries de grandeurs variables et d'obtenir ainsi une estimation combinée de la fréquence de caractère p , on entrevoit, par ce mode de représentation, une possibilité de résolution de l'équation du Likelihood. L'estimation par le Maximum Likelihood est la solution $\hat{p}_s$ de l'équation $D_s(p) = 0$, et la variance de $\hat{p}_s$ est donnée par l'inverse de W_s pour la valeur $\hat{p}_s$. Aussi, pour obtenir les $\hat{p}_s$, peut-on calculer une table des fonctions $[E_s(r) - \delta(k)] / pq$ et des poids w_s pour des valeurs p appropriées, distantes de 0,05 les unes des autres par exemple; le calcul de D_s à partir de valeurs de la table voisines de la solution $\hat{p}_s$ cherchée, et l'interpolation inverse de ces D_s fournissent, en général, une solution suffisamment bonne de $\hat{p}_s$. On déterminera de même W_s pour différentes valeurs de part et d'autre de l'estimation $\hat{p}_s$; l'interpolation directe pour la valeur $\hat{p}_s$ donne alors l'inverse de la variance de l'estimation par le Maximum Likelihood.

Comme d'après (3.25) $V(D_s) = V\left(\frac{\partial L_s}{\partial p}\right)$ est approximativement égal à:

$$= -\Delta \frac{\partial L_s}{\partial p} / \Delta p = -\Delta D_s / \Delta p = W_s \quad (3.36)$$

et comme
$$V(\hat{p}_s) = \frac{1}{V\left(\frac{\partial L_s}{\partial p}\right)},$$

on obtient une valeur approchée de la variance de $\hat{p}_s$, en divisant la variation négative $-\Delta p$ de p par la variation correspondante ΔD_s du score D_s . Cette possibilité de calculer la variance est une des raisons pour lesquelles le facteur $1/pq$, qui semblait tout d'abord superflu, a été conservé dans l'expression de D_s .

Pour les applications en génétique, on peut se contenter d'une interpolation linéaire dans la plupart des cas, et il suffit ainsi de déterminer D_s et W_s en partant des valeurs de la table voisines, de part et d'autre, de $\hat{p}_s$.

Les scores pondérés de Finney.

Elaborée par R.A. FISHER et utilisée pour déterminer le linkage de deux caractères, la théorie des scores a été développée particulièrement par FINNEY (1940, 1947/49). La théorie des scores est étroitement liée à la résolution de l'équation du Maximum Likelihood par la méthode d'itération (voir cependant le point de vue plus général de FINNEY, 1949).

Si, dans l'ensemble, on a n_s fratries de grandeur s avec r_s sujets atteints et qu'il faut à nouveau estimer la fréquence p de sujets atteints dans la

répartition binomiale tronquée, le procédé d'estimation consiste à tirer $\hat{p}_s$ soit de l'équation qui résulte de l'identification de l'observation et de l'espérance mathématique, soit de l'équation du Maximum Likelihood $\partial L_s / \partial p = D_s = 0$ admettant les mêmes solutions (non triviales). Pour la résolution de l'équation, on se sert du procédé d'approximation de NEWTON, qui garantit en même temps la convergence des approximations. Si $\hat{p}_s$ est la solution de l'équation du Maximum Likelihood, et p_0 une approximation quelconque, on aura, comme on sait

$$D_s(\hat{p}_s) = D_s(p_0) + (\hat{p}_s - p_0) \frac{\partial D_s}{\partial p}(p_0^*) = 0, \quad p_0 < p_0^* < \hat{p}_s \quad (3.37)$$

Si l'on calcule $\partial D_s / \partial p$ pour la valeur p_0 au lieu de p_0^* , le membre droit de l'équation ne sera plus tout à fait valable, à moins que dans $(\hat{p}_s - p_0)$ on ne substitue à $\hat{p}_s$ la valeur p_1 qui est, comme on peut le démontrer, une meilleure approximation. La résolution par rapport à p_1 donne, si l'on tient encore compte de

$$\begin{aligned} \frac{\partial D_s}{\partial p}(p_0) &= -W_s(p_0), \\ p_1 &= p_0 + \frac{D_s(p_0)}{W_s(p_0)}. \end{aligned} \quad (3.38)$$

De son côté, l'approximation p_1 peut être utilisée pour le calcul d'une nouvelle approximation p_2 etc.

Puisque D_s et W_s contiennent des puissances de q telles que q^s , le procédé d'approximation, en dépit d'une convergence rapide, nécessite des calculs fastidieux. On s'arrêtera au premier échelon d'approximation, et on définira comme fonction du paramètre inconnu p le score (FINNEY, 1950)

$$Y_s = p + \frac{D_s(p)}{W_s(p)}. \quad (3.39)$$

Ce score Y_s dépend non seulement de p , mais en plus, par l'intermédiaire de D_s , du nombre total r_s de sujets atteints dans les n_s fratries. Comme pour les D -scores, on peut définir un score y_s pour la contribution d'une seule fratrie :

$$y_s = p + \frac{d_s(p)}{w_s(p)}. \quad (3.40)$$

Si la probabilité de base est p , et qu'on tienne compte de la relation

$$E(d_s) = 0 \text{ et } V(d_s) = w_s,$$

on trouve immédiatement

$$E(y_s) = p + \frac{E(d_s)}{w_s(p)} = p \quad (3.41)$$

de même que

$$V(y_s) = \frac{V(d_s)}{w_s^2(p)} = \frac{w_s(p)}{w_s^2(p)} = \frac{1}{w_s(p)} \quad (3.42)$$

et enfin $E(Y_s) = p$, $V(Y_s) = \frac{1}{W_s(p)}$. (3.43)

Ces scores ont une propriété en commun avec la fonction d'estimation non biaisée: pour des échantillons d'un ensemble de base où la fréquence de caractère est p , leur espérance mathématique coïncide avec p . Ces scores dépendent non seulement des observations r ou r_s , mais du paramètre à déterminer. Si l'on choisit ce dernier de façon à ce que le score coïncide avec son espérance mathématique, on obtiendra pour n_s fratries de grandeur s immédiatement l'équation du Maximum Likelihood:

$$D_s(\hat{p}_s) = 0,$$

la variance $V(Y_s)$ du score pour la valeur $\hat{p}_s$ est celle de $\hat{p}_s$, donc de $1/W_s(\hat{p}_s)$.

En vue de l'application pratique, en particulier pour le groupement des fratries de différentes grandeurs, il est plus indiqué de considérer le score pondéré

$$w_s y_s = w_s p + d_s(p) \quad (3.44)$$

car les grandeurs d_s et D_s sont additives, tandis que d_s/w_s et D_s/W_s ne le sont pas. L'addition des n_s quantités relatives aux fratries de grandeur s donne, compte tenu de $n_s w_s = W_s$ et $\sum_1^{n_s} d_s = D_s$,

$$W_s Y_s = W_s p + D_s(p), \quad (3.45)$$

où Y_s est lié à y_s par la relation $Y_s = \sum y_s / n_s$. L'estimation de p au moyen de scores pondérés, dans le cas de fratries de différentes grandeurs, sera traitée plus tard.

D'après les relations (3.44) et (3.45), le score pondéré $w_s y_s$ ou $W_s Y_s$ n'est autre que le score d_s ou D_s augmenté de la valeur $w_s p$ ou $W_s p$. Si, dans le membre droit de l'équation, on fait prendre à p différentes valeurs, le score D_s disparaîtra pour une valeur déterminée $\hat{p}_s$, et il restera $W_s(\hat{p}_s) \hat{p}_s$. Pour obtenir $\hat{p}_s$, on n'a qu'à diviser le score pondéré relatif à la valeur $\hat{p}_s$ par W_s calculé pour cette même valeur $\hat{p}_s$. Comme $\hat{p}_s$ n'est pas connu d'avance, on choisira des valeurs appropriées supérieures et inférieures p' et p'' , avec $p' < \hat{p}_s < p''$. Les scores pondérés correspondant à p' et p'' , divisés par les poids W'_s et W''_s pris aux mêmes valeurs, donneront directement deux estimations améliorées $\hat{p}' > p'$ et $\hat{p}'' < p''$. Par interpolation linéaires, ces dernières permettront d'améliorer l'approximation de $\hat{p}_s$ encore davantage s'il est nécessaire. De même, l'interpolation linéaire à partir des poids W'_s et W''_s fournira le poids relatif à la valeur interpolée de p , lequel poids déterminera à son tour la variance de $\hat{p}_s$.

Le calcul des scores est facilité par des tables indiquant, pour des valeurs de p suffisamment rapprochées, les poids $w_s(p)$ ainsi que les valeurs d'une autre fonction $b_s(p)$ entrant dans le calcul et ne dépendant que de p . Pour trouver $b_s(p)$, on substitue dans le score pondéré $w_s y_s$ pour chaque fratrie le score d_s suivant les formules (3.30) - (3.33).

$$\text{On trouve } w_s y_s = p w_s + \left[\frac{r \cdot \delta(k)}{pq} - \frac{E_s(r) \cdot \delta(k)}{pq} \right]$$

d'où l'on écrit

$$\begin{aligned} w_s y_s &= \frac{r \cdot \delta(k)}{pq} - \left[\frac{E_s(r) \cdot \delta(k)}{pq} - p w_s \right] \\ &= \frac{r \cdot \delta(k)}{pq} - b_s(p) \end{aligned} \quad (3.46)$$

Ci-dessous, on a réuni les formules des b_s se rapportant aux différents procédés de sélection et complétées des formules correspondantes des w_s :

$$\rho = 1, k = 0 : b_s = 0$$

$$w_s = \frac{s-1}{pq} \quad (3.47)$$

$$\rho = 1, k = 1 : b_s = \frac{s^2 pq^{s-2}}{(1-q^s)^2}$$

$$w_s = \frac{s}{pq} \frac{1-q^s - spq^{s-1}}{(1-q^s)^2} \quad (3.48)$$

$$\rho = 2, k = 0 : b_s = \frac{(s-1)^2 pq^{s-3}}{(1-q^{s-1})^2}$$

$$w_s = \frac{s-1}{pq} \frac{1-q^{s-1} - (s-1) pq^{s-2}}{(1-q^{s-1})^2} \quad (3.49)$$

$$\rho = 2, k = 1 : b_s = \frac{s(s-1) pq^{s-3} (sp-1 + q^s)}{(1-q^s - spq^{s-1})^2}$$

$$w_s = \frac{s}{pq} \frac{(1-q^{s-1})^2 - (s-1)^2 p^2 q^{s-2}}{(1-q^s - spq^{s-1})^2} \quad (3.50)$$

FINNEY (1950) a calculé des tables où sont données, sous la désignation de B_s et W_s , les valeurs de b_s et w_s dans le cas où $\rho = 1$, $k = 1$, pour $p = 0,01$ (0,01) 0,05 (0,05) 0,95 et pour $s = 2$ (1) 20. Les tables Ia - Id jointes au présent travail en donnent des extraits pour $P = 0,05$ (0,05) 0,95 et des valeurs complémentaires pour le cas où $\rho = 2$, $k = 1$ et le domaine des valeurs $p = 0,05$ (0,05) 0,95, $s = 3$ (1) 16. On n'a pas besoin de tables pour le cas où $\rho = 1$, $k = 0$, et on peut ramener le cas où $\rho = 2$, $k = 0$ à celui où $\rho = 1$, $k = 1$, en prenant pour un s déterminé la valeur de la table correspondant à $s-1$.

La précision de l'estimation, obtenue par l'utilisation des tables Ia - Id suivie d'une interpolation linéaire, conformément au procédé décrit plus haut, a, de toute évidence, certaines limites. Si l'on exige des valeurs plus exactes, le score pondéré peut être calculé non seulement pour les deux valeurs de la table voisines de $\hat{p}_s$, mais pour plusieurs autres valeurs de la table, l'interpolation inverse étant ensuite appliquée. Ou bien, on peut passer d'une approximation à d'autres grâce à un procédé d'itération, mais alors on ne pourra plus tirer les grandeurs b_s et w_s des tables.

Pour terminer, nous voulons encore exprimer les scores d_s en fonction des scores y_s . D'après (3.40), on a

$$d_s(p) = w_s(y_s - p) = \left[\frac{r - \delta(k)}{pq} - p \right] - b_s(p). \quad (3.51)$$

Comme non seulement les valeurs de $w_s(p)$, mais encore celles de $b_s(p)$ ont été réunies dans des tables pour des valeurs appropriées de p , on peut aussi calculer les scores $d_s(p)$ sans peine.

IV. LES PROCÉDÉS D'ESTIMATION DE WEINBERG, RIDER ET MOORE.

Les procédés d'estimation de WEINBERG, RIDER et MOORE partent, dans une large mesure, d'autres hypothèses que le procédé du Maximum Likelihood. Par opposition à la méthode du Maximum Likelihood, où la fonction d'estimation est habituellement obtenue comme solution d'une équation transcendente dans laquelle on a fait entrer les valeurs observées, la fonction d'estimation est, pour les procédés d'estimation décrits dans la suite, une fonction rationnelle plus ou moins simple des observations.

Si l'on se limite, pour l'instant, aux fratries de grandeur s , les estimations $\hat{p}_s$ ont la forme :

$$\hat{p}_s = \frac{F_s}{G_s}, \quad (4.1)$$

où F_s , G_s , sont des sommes de fonctions rationnelles entières ou fractionnaires du nombre de sujets atteints, la somme étant étendue aux n_s fratries de grandeur s ; donc

$$F_s = f_1(r) + \dots + f_{n_s}(r), \quad (4.2)$$

$$G_s = g_1(r) + \dots + g_{n_s}(r). \quad (4.3)$$

On entend par $f_i(r)$ ou $g_i(r)$ la valeur de la fonction f ou g pour le nombre r de sujets atteints observés dans la i -ième fratrie. Puisque les fonctions f ou g sont identiques et leurs arguments mutuellement indépendants, on trouve

$$E(F_s) = n_s E_s f(r) = n_s \sum_{r=\rho}^s f(r) P_{sr} \quad (4.4)$$

$$E(G_s) = n_s E_s g(r) = n_s \sum_{r=\rho}^s g(r) P_{sr} \quad (4.5)$$

Il faut comprendre les espérances mathématiques à gauche dans le sens d'une répartition commune des n_s valeurs r , les espérances mathématiques à droite dans le sens d'une répartition du nombre de sujets atteints dans une seule fratrie. P_{sr} sont les probabilités définies dans (2.3) - (2.6). La somme s'étend de $\rho = 1$, ou $\rho = 2$ à s , suivant qu'il s'agit de la répartition binomiale privée des classes $r = 0$ et $r = 1$.

On peut également montrer que

$$V(F_s) = n_s V_s[f(r)] = n_s \left[\sum_{r=\rho}^s f^2(r) P_{sr} - \left(\sum_{r=\rho}^s f(r) P_{sr} \right)^2 \right] \quad (4.6)$$

$$V(G_s) = n_s V_s[g(r)] = n_s \left[\sum_{r=\rho}^s g^2(r) P_{sr} - \left(\sum_{r=\rho}^s g(r) P_{sr} \right)^2 \right] \quad (4.7)$$

et

$$\begin{aligned} \text{Cov}(F_s, G_s) = n_s \text{Cov}_s[f(r), g(r)] = n_s \left[\sum_{r=\rho}^s f(r) g(r) P_{sr} - \right. \\ \left. \left(\sum_{r=\rho}^s f(r) P_{sr} \right) \left(\sum_{r=\rho}^s g(r) P_{sr} \right) \right] \end{aligned} \quad (4.8)$$

Remarquons d'emblée que les procédés d'estimation de WEINBERG, RIDER et MOORE s'appuient sur le fait que, pour un choix approprié des fonctions $f(r)$ et $g(r)$, on a

$$\frac{E(F_s)}{E(G_s)} = p, \quad (4.9)$$

En d'autres termes, le quotient des espérances mathématiques, au numérateur et au dénominateur, est égal à la valeur du paramètre dans l'ensemble de base. D'une façon générale, cela ne signifie pas encore que l'espérance mathématique du quotient

$$F_s / G_s = \hat{p}_s$$

soit égale à p . En général, $E(\hat{p}_s)$ différera de p d'une grandeur dépendant de n_s , de p et du genre des fonctions f et g choisies. En supposant que les moments simples et mixtes du troisième degré de F et G soient petits par rapport à $V(F)$, $\text{Cov}(F, G)$, $V(G)$ qui de leur côté doivent être petits, on peut déterminer approximativement la grandeur de ce biais selon la formule

$$E\left(\frac{F}{G}\right) - \frac{E(F)}{E(G)} \approx \frac{1}{[E(G)]^2} \left[\frac{E(F)}{E(G)} V(G) - \text{Cov}(F, G) \right], \quad (4.10)$$

(cf. v. MISES (1931)). Puisque, par la suite, $E(F) / E(G)$ est toujours égal à p , on peut écrire

$$E\left(\frac{F}{G}\right) - \frac{E(F)}{E(G)} \approx \frac{1}{[E(G)]^2} [pV(G) - \text{Cov}(F, G)], \quad (4.11)$$

ou encore

$$E \left(\frac{F}{G} \right) - \frac{E(F)}{E(G)} \approx \frac{1}{E^2(G)} [pE(G^2) - E(FG)] \quad (4.12)$$

De même, on ne peut indiquer qu'approximativement, en général, la variance du quotient $\hat{p}_s = F_s / G_s$. D'après (4.10), on a (voir par exemple v. MISES, 1931):

$$V \left(\frac{F}{G} \right) \approx \frac{1}{[E(G)]^2} [V(F) - 2 \frac{E(F)}{E(G)} \text{Cov}(F, G) + (\frac{E(F)}{E(G)})^2 V(G)], \quad (4.13)$$

ou bien

$$V \left(\frac{F}{G} \right) \approx \frac{1}{E^2(G)} [V(F) - 2p \text{Cov}(F, G) + p^2 V(G)], \quad (4.14)$$

ou encore

$$V \left(\frac{F}{G} \right) \approx \frac{1}{E^2(G)} [E(F^2) - 2p E(F, G) + p^2 E(G^2)]. \quad (4.15)$$

Les fondements des méthodes de WEINBERG, RIDER et MOORE seront exposés par la suite. Puis on indiquera la variance de $\hat{p}_s$ ainsi que le biais de $\hat{p}_s$ par rapport à la valeur p du paramètre dans l'ensemble de base. De nouveau, on commencera par ne considérer que des fratries d'une grandeur déterminée s .

LE PROCÉDÉ DE WEINBERG

(Méthode par réduction $k = 0$ et «Geschwistermethode» $k = 1$).

Le procédé de WEINBERG, de par sa simplicité, se trouve être répandu assez largement dans la littérature de la génétique médicale. WEINBERG, déjà en 1912, semble avoir remarqué que le nombre relatif moyen $\bar{i}/s$ des tarés dans des fratries de grandeur s ne fournit pas une estimation utilisable de la fréquence de caractère p dans l'ensemble de base, si le procédé de sélection ne permet d'obtenir qu'une partie des fratries ayant des parents d'un génotype déterminé. Les caractères récessifs rares ($P = 0,25$) en donnent un exemple typique, car dans ce cas on ne peut tenir compte des fratries avec les deux parents hétérozygotes pour le caractère en question que si elles présentent au minimum un sujet atteint. Partant de l'idée que la constitution génétique d'un enfant n'est déterminée que par celle de ses parents et non pas par celle de ses frères et sœurs, WEINBERG ne considère tout d'abord que les germains d'un sujet atteint, les germains du proband par exemple, puis il forme le quotient mentionné. Ceci lui fournit alors une estimation utilisable de la fréquence de caractère, à supposer que la fratrie présente un seul

proband. En présence de plusieurs probands, on introduira leur nombre comme poids (méthode de probands que R. A. FISHER a désigné par « sib method ») ; si l'on suppose, en particulier, que chacun des r sujets atteints est proband (« Geschwistermethode »), ce poids est égal à r .

Supposons pour commencer qu'on ne considérera que des fratries avec au moins un sujet atteint ($\rho = 1$) ; pour une fratrie unique de grandeur s (≥ 2) et dans le cas de la sélection simple ($k = 0$), les fonctions $f(r)$ et $g(r)$ sont donc :

$$k = 0 : f(r) = r - 1, g(r) = s - 1, s \geq 2, r \geq 1 \quad (4.16)$$

Il s'ensuit de (2.8) que

$$E_s [f(r)] = (s - 1) p, E_s [g(r)] = s - 1 \quad (4.17)$$

De plus

$$F_s = \sum_1^{(n_s)} (r - 1) = r_s \cdot n_s, G_s = \sum_1^{(n_s)} (s - 1) = s n_s \cdot n_s \quad (4.18)$$

et

$$\hat{p}_s = \frac{F_s}{G_s} = \frac{\sum (r - 1)}{\sum (s - 1)} = \frac{r_s \cdot n_s}{n_s (s - 1)}. \quad (4.19)$$

L'application de (4.17) donne immédiatement

$$\frac{E(F_s)}{E(G_s)} = p. \quad (4.20)$$

En utilisant les fréquences n_{sr} , on peut aussi écrire les sommes $\sum_1^{(n_s)} (...)$ sous la forme

$$\sum_{r=\rho}^s (...) n_{sr}$$

$$k = 1 : f(r) = r(r - 1), g(r) = r(s - 1), s \geq 2, r \geq 1 \quad (4.21)$$

On trouve

$$E_s (f(r)) = \frac{s(s - 1)p^2}{1 - q^s}, E_s (g(r)) = \frac{s(s - 1)p}{1 - q^s} \quad (4.22)$$

De plus,

$$F_s = \sum_1^{(n_s)} r(r - 1), G_s = \sum_1^{(n_s)} r(s - 1) \quad (4.23)$$

et

$$\hat{p}_s = \frac{F_s}{G_s} = \frac{\sum r(r - 1)}{\sum r(s - 1)}. \quad (4.24)$$

On a à nouveau

$$\frac{E(F_s)}{E(G_s)} = p. \quad (4.25)$$

Lors de l'extension de la méthode de WEINBERG au cas de la répartition binomiale privée des classes $r = 0$ et $r = 1$ ($\rho = 2$), on est amené à penser que, comme dans le cas de la méthode du Maximum Likelihood, l'estimation $\hat{p}_s$ pour $k = 0$ devrait correspondre à celle où $\rho = 1$, $k = 1$, pourvu que, dans l'estimation, on remplace partout s et r par $s-1$ et $r-1$ respectivement (les frères et sœurs du proband dans la répartition privée des classes $r = 0$, et $r = 1$, forment une répartition privée de la classe $r = 0$). Pour $\rho = 2$ et $k = 0$, on exige donc que

$$f(r) = (r-1)(r-2), \quad g(r) = (r-1)(s-2), \quad s \geq 3, \quad r \geq 2. \quad (4.26)$$

Il s'ensuit que

$$E_s(f(r)) = \frac{(s-1)(s-2)p^2}{1-q^{s-1}}, \quad E_s(g(r)) = \frac{(s-1)(s-2)p}{1-q^{s-1}} \quad (4.27)$$

et

$$F_s = \sum_1^{(n_s)} (r-1)(r-2), \quad G_s = \sum_1^{(n_s)} (r-1)(s-2) \quad (4.28)$$

$$\hat{p}_s = \frac{F_s}{G_s} = \frac{\sum (r-1)(r-2)}{\sum (r-1)(s-2)}. \quad (4.29)$$

On aura à nouveau

$$\frac{E(F_s)}{E(G_s)} = p. \quad (4.30)$$

Du point de vue purement formel, il est alors naturel de poser, dans le cas où $k = 1$:

$$k = 1 : f(r) = r(r-1)(r-2), \quad g(r) = r(r-1)(s-2), \quad s \geq 3, \quad r \geq 2 \quad (4.31)$$

$$E_s(f(r)) = \frac{s(s-1)(s-2)p^3}{1-q^s - spq^{s-1}}, \quad E_s(g(r)) = \frac{s(s-1)(s-2)p^2}{1-q^s - spq^{s-1}} \quad (4.32)$$

$$F_s = \sum_1^{(n_s)} r(r-1)(r-2), \quad G_s = \sum_1^{(n_s)} r(r-1)(s-2), \quad s \geq 3, \quad r \geq 2 \quad (4.33)$$

$$\hat{p}_s = \frac{F_s}{G_s} = \frac{\sum r(r-1)(r-2)}{\sum r(r-1)(s-2)}, \quad (4.34)$$

$$\frac{E(F_s)}{E(G_s)} = p. \quad (4.35)$$

La relation entre les expressions dans lesquelles $k = 0$ et $k = 1$, peut être trouvée dans les formules précédentes comme dans le calcul du biais et de

la variance ci-dessous en considérant $k = 0$ et $k = 1$ comme cas limites de la méthode par probands selon WEINBERG avec k comme degré de sélection général et un nombre t de probands comme variable aléatoire supplémentaire. C'est ainsi qu'on peut relier entre eux les deux cas limites $k = 0$ et $k = 1$, comme dans la méthode du Maximum Likelihood.

Pour calculer le biais et la variance de l'estimation $\hat{p}_s$ à l'aide des formules (4.12) et (4.15), on a besoin de grandeurs $E(F^2) = n_s E(f^2)$. Lorsqu'on suppose que la fréquence de caractère, dans l'ensemble de base, est égale à p , on obtient :

$$\rho = 1, k = 0 : E_s [f^2(r)] = (s-1) [p + (s-2)p^2], \quad (4.36)$$

$$E_s [f(r) g(r)] = (s-1)^2 p, \quad (4.37)$$

$$E_s [g^2(r)] = (s-1)^2, \quad (4.38)$$

$$\rho = 1, k = 1 : E_s [f^2(r)] = \frac{s(s-1)p}{1-q^s} [(s-2)(s-3)p^3 + 4(s-2)p^2 + 2p] \quad (4.39)$$

$$E_s [f(r) g(r)] = \frac{s(s-1)p}{1-q^s} [(s-1)(s-2)p^2 + 2(s-1)p] \quad (4.40)$$

$$E_s [g^2(r)] = \frac{s(s-1)p}{1-q^s} [(s-1)^2 p + (s-1)] \quad (4.41)$$

$$\rho = 2, k = 0 : E_s [f^2(r)] = \frac{(s-1)(s-2)p}{1-q^{s-1}} [(s-3)(s-4)p^3 + 4(s-3)p^2 + 2p] \quad (4.42)$$

$$E_s [f(r) g(r)] = \frac{(s-1)(s-2)p}{1-q^{s-1}} [(s-2)(s-3)p^2 + 2(s-2)p] \quad (4.43)$$

$$E_s [g^2(r)] = \frac{(s-1)(s-2)p}{1-q^{s-1}} [(s-2)^2 p + (s-2)] \quad (4.44)$$

$$\rho = 2, k = 1 : E_s [f^2(r)] = \frac{s(s-1)(s-2)p^2}{1-q^s - spq^{s-1}} [(s-3)(s-4)(s-5)p^4 + 9(s-3)(s-4)p^3 + 18(s-3)p^2 + 6p] \quad (4.45)$$

$$E_s [f(r) g(r)] = \frac{s(s-1)(s-2)p^2}{1-q^s - spq^{s-1}} [(s-2)(s-3)(s-4)p^3 + 6(s-2)(s-3)p^2 + 6(s-2)p] \quad (4.46)$$

$$E_s [g^2(r)] = \frac{s(s-1)(s-2)p^2}{1-q^s - spq^{s-1}} [(s-2)^2(s-3)p^2 + 4(s-2)^2 p + 2(s-2)] \quad (4.47)$$

On trouve immédiatement ces espérances mathématiques, si l'on écrit les fonctions $f^2(r)$, $f(r)g(r)$, $g^2(r)$ sous la forme

$$\sum_i a_i (r-1)^{[i]} \text{ pour } k=0, \text{ soit sous la forme } \sum_i b_i r^{[i]} \text{ lorsque } k=1, \text{ avec}$$

$$(r-1)^{[0]} = r^{[0]} = 1 \text{ et } (r-1)^{[i]} = (r-1)(r-2) \dots (r-i)$$

$$r^{[i]} = r(r-1)(r-2) \dots (r-i+1) \quad \text{pour } i > 0$$

Par exemple pour $\rho = 1$, $k = 1$:

$$[r(r-1)]^2 = r(r-1)(r-2)(r-3) + 4r(r-1)(r-2) + 2r(r-1),$$

ce qui nous amène sans peine à l'expression donnée ci-dessus.

L'application de la formule (4.12) conduit aux valeurs suivantes de l'espérance mathématique $E(\hat{\rho}_s)$

$$\rho = 1, k = 0: E(\hat{\rho}_s) = 0 \quad (4.48)$$

$$\rho = 1, k = 1: E(\hat{\rho}_s) \approx p \cdot q \frac{1-q^s}{sn_s} \approx p \cdot \frac{s-1}{E(G_s)} pq \quad (4.49)$$

$$\rho = 2, k = 0: E(\hat{\rho}_s) \approx p \cdot q \frac{1-q^{s-1}}{(s-1)n_s} \approx p \cdot \frac{s-2}{E(G_s)} pq \quad (4.50)$$

$$\rho = 2, k = 1: E(\hat{\rho}_s) \approx p \cdot q \frac{1-q^s - spq^{s-1}}{n_s s (s-1) p} [4 + 2(s-3)p]$$

$$\approx p \cdot \frac{2(s-2) [2 + (s-3)p]}{E(G_s)} pq \quad (4.51)$$

De même, on aura d'après (4.15) pour la variance de l'estimation $\hat{\rho}_s$

$$\rho = 1, k = 0: V(\hat{\rho}_s) = \frac{pq}{n_s (s-1)} = \frac{pq}{E(G_s)} \quad (4.52)$$

$$\rho = 1, k = 1: V(\hat{\rho}_s) \approx pq \frac{(1-q^s) [2 + (s-3)p]}{n_s s (s-1) p} \approx pq \frac{2 + (s-3)p}{E(G_s)} \quad (4.53)$$

$$\rho = 2, k = 0: V(\hat{\rho}_s) \approx pq \frac{(1-q^{s-1}) [2 + (s-4)p]}{n_s (s-1)(s-2)p} \approx pq \frac{2 + (s-4)p}{E(G_s)} \quad (4.54)$$

$$\rho = 2, k = 1: V(\hat{\rho}_s) \approx pq \frac{1-q^s - spq^{s-1}}{n_s s (s-1)(s-2)p^2} [(s-3)(s-8)p^2 + 4(2s-7)p + 6]$$

$$\approx pq \frac{6 + 4(2s-7)p + (s-3)(s-8)p^2}{E(G_s)} \quad (4.55)$$

Il est nécessaire de faire quelques remarques au sujet des espérances mathématiques et des variances de l'estimation $\hat{p}_s$, indiquées ci-dessus et calculées toujours pour des fratries d'une grandeur déterminée s :

1) Un cas particulièrement intéressant est celui où $\rho = 1$, $k = 0$, c'est-à-dire sélection simple avec répartition binomiale privée de la classe $r = 0$. Les résultats coïncident exactement avec ceux qu'on avait obtenus pour la méthode de réduction selon le procédé du Maximum Likelihood. Là aussi, le cas de $\rho = 1$, $k = 0$ est le seul où l'estimation soit non biaisée, c'est-à-dire satisfaisant exactement la relation $E(\hat{p}_s) = p$. Ceci provient du fait que le dénominateur $G_s = n_s(s-1)$ ne contient aucune variable aléatoire, mais uniquement des constantes. La variance de $\hat{p}_s$ est, comme il fallait s'y attendre, $pq / n_s(s-1)$, c'est-à-dire qu'elle correspond exactement à la variance binomiale.

2) Il faut calculer, en principe, les espérances mathématiques et les variances pour la valeur $\hat{p}_s$, sinon pour la valeur déterminée par une certaine hypothèse héréditaire, lorsqu'il s'agit de vérifier cette dernière. Sans risque d'erreur trop grand, on peut remplacer dans le dénominateur des formules (4.48) - (4.55) $E(G_s)$ par G_s lorsque n_s est grand, ce qui simplifie le calcul de ces expressions.

3) Comme le montrent les formules de l'espérance mathématique de $\hat{p}_s$, les estimations suivant la méthode de WEINBERG sont correctes («consistent»), mais cependant biaisées, sauf dans le cas où $\rho = 1$, $k = 0$. A l'exception de ce même cas $\rho = 1$, $k = 0$, la variance asymptotique ($n_s \rightarrow \infty$) dans une seule fratrie est en général plus grande que la variance correspondante obtenue suivant la méthode du Maximum Likelihood. Par conséquent, la valeur inverse w_s de cette variance est inférieure à la quantité correspondante déduite du Maximum Likelihood (voir table IIa - IIb). En général, l'efficacité relative J_{WB} reste inférieure à 1 (voir table V et diagramme). Les estimations d'après WEINBERG ne sont donc pas efficaces («efficient»). De plus, elles ne sont pas exhaustives («sufficient»).

4) Dans les cas où $\rho = 1$, $k = 1$ («Geschwistermethode»), v. BEHR a donné une évaluation plus exacte de $E(\hat{p}_s)$, donc du biais, en utilisant la répartition simultanée du nombre r des sujets atteints dans les n_s fratries. Il trouve

$$E(\hat{p}_s) = p \cdot q \frac{1 - q^s}{s(n_s + 1)} F\left(\frac{1}{s} + 1, 1, n_s + 2, 1 - q^s\right) \quad (4.56)$$

$$= p \cdot \frac{q}{n_s + 1} \cdot \frac{1 - q^s}{s} + 0 \left(\frac{1}{n_s^2}\right) \quad (4.57)$$

où $F(\)$ représente la fonction hypergéométrique. Ainsi que BERWALD (1924), il trouve la variance sous la forme

$$V(\hat{p}_s) = \frac{1 - q^s}{s(n_s + 1)} \left[pq + \frac{2q^2}{s-1} \right] + O\left(\frac{1}{n_s^2}\right). \quad (4.58)$$

Au facteur $1/n_s$ près, cette forme concorde avec l'expression (4.53) ci-dessus, dans laquelle ce facteur valait $1/(n_s + 1)$. Ici aussi, on peut déduire les deux cas limites $k = 0$ et $k = 1$ d'une relation plus générale, correspondant à la méthode par proband. Pour $\rho = 2$, $k = 1$, c'est-à-dire pour la répartition privée des classes $r = 0$ et $r = 1$, on ne connaît pas de développements asymptotiques simples.

5) On traitera plus tard la méthode de WEINBERG pour l'estimation de p à partir de fratries de grandeurs différentes s .

6) Dans le cas de la méthode par proband, on peut estimer le degré de sélection k à partir du nombre de probands et de sujets atteints en appliquant le même principe que pour l'estimation de p .

LE PROCÉDÉ DE RIDER

Le mode d'estimer la fréquence de caractère p dans des fratries de grandeur s suivant la méthode de WEINBERG, peut être justifié par des considérations d'ordre génétique pour le moins dans le cas où $\rho = 1$, $k = 0$ ou 1 ; tel n'est plus le cas pour les méthodes de RIDER et de MOORE qui vont être traitées. Ces méthodes au contraire s'obtiennent immédiatement en considérant une répartition binomiale dans laquelle une ou plusieurs classes en l'occurrence voisines de zéro ont été laissées de côté. C'est ainsi que disparaît d'emblée la relation entre les deux cas limites, à savoir celui de la sélection complète ($k = 1$) et celui de la sélection simple ($k = 0$), relation telle qu'elle existait dans la méthode du Maximum Likelihood ainsi que dans celle de WEINBERG. En effet, les considérations et les formules qui vont suivre se rapportent au cas de la sélection complète ($k = 1$), même si nous ne l'indiquons plus explicitement. On peut toujours passer au cas de la sélection simple ($k = 0$), en considérant la fratrie réduite de son unique proband. Mais le rapport entre les formules pour $k = 1$ et $k = 0$ cesse d'exister, et il est nécessaire, dans chaque cas, de définir directement $\hat{p}_s$ pour $k = 0$.

Dans la justification du procédé de RIDER, nous nous limiterons à la répartition binomiale privée de la classe $r = 0$ ou des classes $r = 0$ et $r = 1$, bien que les relations puissent être étendues au cas d'un nombre quelconque de classe supprimées.

Soit n_{sr} le nombre de fratries (parmi les n_s fratries) qui présentent r sujets atteints ($r \geq \rho$; $\rho = 1$ pour la répartition binomiale privée de la classe $r = 0$; $\rho = 2$ pour la répartition privée des classes $r = 0$ et $r = 1$); on a

$$n_s = \sum_{r=\rho}^s n_{sr}, \quad r_s = \sum_{r=\rho}^s r n_{sr}, \quad v_s = \sum_{r=\rho}^s r^2 n_{sr} \quad (4.59)$$

De plus, on désignera par N_s le nombre total inconnu de fratries de grandeur s , y compris les fratries tronquées. En supposant un p indéterminé, on peut substituer aux grandeurs n_s, r_s, v_s les expressions

$$n'_s = n_s + N_s \sum_{r=0}^{\rho=1} \binom{s}{r} p^r q^{s-r} \quad (4.60)$$

$$r'_s = r_s + N_s \sum_{r=0}^{\rho=1} r \binom{s}{r} p^r q^{s-r} \quad (4.61)$$

$$v'_s = v_s + N_s \sum_{r=0}^{\rho=1} r^2 \binom{s}{r} p^r q^{s-r} \quad (4.62)$$

Le quotient r'_s / n'_s est une estimation du premier moment sp de la répartition non tronquée, le quotient v'_s / r'_s une estimation de $[spq + (sp)^2] / sp = q + sp$. On pose alors

$$r'_s = sp \cdot n'_s, \quad v'_s = (q + sp) \cdot r'_s. \quad (4.63)$$

L'élimination de N_s , dont on n'aura plus besoin, donne une relation qui permet d'estimer p .

Dans le cas où $\rho = 1$, les trois relations (4.60) - (4.62) sont :

$$n'_s = n_s + N_s q^s, \quad r'_s = r_s, \quad v'_s = v_s.$$

De la deuxième équation (4.63), il résulte

$$\begin{aligned} \frac{v_s}{r_s} &= q + sp = 1 + (s-1)p \\ \text{et de là} \quad \hat{p}_s &= \frac{v_s - r_s}{r_s (s-1)} = \frac{\sum_{r=1}^s n_{sr} r (r-1)}{\sum_{r=1}^s n_{sr} r (s-1)} \end{aligned} \quad (4.64)$$

Ceci correspond entièrement à l'estimation selon WEINBERG (formule 4.24).

Dans le cas où $\rho = 2$, les trois relations (4.60) - (4.62) sont

$$n'_s = n_s + N_s q^s + N_s spq^{s-1} = n_s + N_s q^{s-1} (q + sp)$$

$$r'_s = r_s + N_s spq^{s-1} \quad \text{et} \quad v'_s = v_s + N_s spq^{s-1}$$

Donc suivant la première formule dans (4.63)

$$r_s + N_s spq^{s-1} = spn_s + N_s q^{s-1} (q + sp) sp$$

ou

$$r_s - spn_s = N_s spq^{s-1} [q + sp - 1] = N_s s (s-1) p^2 q^{s-1} \quad (4.65)$$

La seconde formule (4.63) donne

$$v_s + N_s spq^{s-1} = (q + sp) [r_s + N_s spq^{s-1}]$$

ou

$$v_s - (q + sp) r_s = N_s s (s-1) p^2 q^{s-1} \quad (4.66)$$

Donc de (4.65) et (4.66)

$$v_s - (q + sp) r_s = r_s - spn_s$$

D'où l'on tire

$$\hat{p}_s = \frac{v_s - 2 r_s}{(s-1) r_s - sn_s} = \frac{\sum_{r=2}^s n_{sr} r (r-2)}{\sum_{r=2}^s n_{sr} [(s-1) r - s]} \quad (4.67)$$

Dans les deux cas ($\rho = 1$ et 2), l'estimation $\hat{p}_s$ est un quotient de sommes de fonctions identiques, la somme étant étendue aux n_s fratries et les fonctions dépendant de la variable aléatoire r des fratries. Pour le degré de sélection $k = 1$, qui seul entre ici en ligne de compte, les fonctions $f(r)$ et $g(r)$ ont la forme

$$\rho = 1, k = 1 : f(r) = r(r-1), \quad g(r) = r(s-1), \quad r \geq 1 \quad (4.68)$$

et si l'on pose $1 - q^s = \alpha_{1s}$,

$$\begin{aligned} E_s [f(r)] &= \frac{s(s-1)p^2}{1-q^s} = \frac{s(s-1)p^2}{\alpha_{1s}}, \\ E_s [g(r)] &= \frac{s(s-1)p}{1-q^s} = \frac{s(s-1)p}{\alpha_{1s}} \end{aligned} \quad (4.69)$$

De plus,

$$\begin{aligned} F_s &= \sum_{r=1}^{(n_s)} r(r-1) = \sum_{r=1}^s n_{sr} r(r-1) \\ G_s &= \sum_{r=1}^{(n_s)} r(s-1) = \sum_{r=1}^s n_{sr} r(s-1). \end{aligned} \quad (4.70)$$

$$D'où \quad E(F_s) / E(G_s) = p \quad (4.71)$$

$$\rho = 2, k = 1 : f(r) = r(r-2), \quad g(r) = r(s-1) - s, \quad r \geq 2 \quad (4.72)$$

De là, on calcule

$$E_s(f(r)) = sp \frac{[(s-1)p \cdot 1 + q^{s-1}]}{1 - q^s - spq^{s-1}} = sp \frac{\beta_s}{\alpha_{2s}} \quad (4.73)$$

$$E_s(g(r)) = \frac{s[(s-1)p \cdot 1 + q^{s-1}]}{1 - q^s - spq^{s-1}} = s \frac{\beta_s}{\alpha_{2s}},$$

et pour raccourcir

$$1 - q^s - spq^{s-1} = \alpha_{2s}, \quad (s-1)p \cdot 1 + q^{s-1} = \beta_s.$$

Les numérateur et dénominateur du quotient qui fournit l'estimation $\hat{p}_s$ sont

$$F_s = \sum_{r=1}^{(n_s)} r(r-2) = \sum_{\rho=2}^s n_{sr} r(r-2) \quad (4.74)$$

$$G_s = \sum_{r=1}^{(n_s)} [r(s-1) - s] = \sum_{\rho=2}^s n_{sr} [r(s-1) - s]$$

Ainsi qu'on le déduit de (4.73), on a

$$E(F_s) / E(G_s) = p. \quad (4.75)$$

Si l'on désire compléter ces estimations de p par le cas de la sélection simple ($k = 0$), il faut définir $f(r)$ et $g(r)$. Puisque dans le cas où $\rho = 1$, $k = 1$, la méthode de RIDER donne le même résultat que celle de WEINBERG, on utilisera, dans le cas où $\rho = 1$, $k = 0$, le procédé d'estimation de WEINBERG, et l'on posera :

$$\rho = 1, k = 0 : f(r) = r-1, \quad g(r) = s-1, \quad r \geq 1, \quad (4.76)$$

avec des espérances telles que (4.17). Pour $\rho = 2$, $k = 0$, on considère les germains du proband, qui forment une répartition privée de la classe $r = 0$ avec $r-1$ et $s-1$ à la place de r et s . Ici également, on emploiera le procédé de WEINBERG, qui donnera :

$$\rho = 2, k = 0 : f(r) = (r-1)(r-2), \quad g(r) = (r-1)(s-2), \quad r \geq 2 \quad (4.77)$$

Le calcul du biais et de la variance de l'estimation $\hat{p}_s$ dans les deux cas précédents ainsi que dans le cas où $\rho = 1$, $k = 1$, est exactement celui de la méthode de WEINBERG, et c'est pourquoi il ne sera pas répété ici. Pour $\rho = 2$, $k = 1$, on a

$$E_s(f^2(r)) = \frac{1}{\alpha_{2s}} [s(s-1)(s-2)(s-3)p^4 + 2s(s-1)(s-2)p^3 - s(s-1)p^2 + sp - spq^{s-1}] \quad (4.78)$$

$$E_s(f(r)g(r)) = \frac{1}{\alpha_{2s}} [s(s-1)^2(s-2)p^3 - s(s-1)p^2 + sp - spq^{s-1}] \quad (4.79)$$

$$E_s(g^2(r)) = \frac{1}{\alpha_{2s}} [s(s-1)^3p^2 - (s+1)s(s-1)p + s^2 - s(p+sp)q^{s-1}] \quad (4.80)$$

Suivant la formule (4.12), on obtient

$$E(\hat{p}_s) \approx p - \frac{1}{n_s} \frac{(s-1)pq}{s\beta_s} \alpha_{2s} \approx p - \frac{(s-1)pq}{E(G_s)} \quad (4.81)$$

Comme auparavant, pour n_s grand, on peut remplacer dans le numérateur $E(G_s)$ par G_s en première approximation; G_s est une grandeur déduite des valeurs de l'échantillon. La variance de $\hat{p}_s$ donne, d'après (4.15):

$$V(\hat{p}_s) \approx \frac{pq}{sn_s} \frac{\alpha_{2s}}{\beta_s^2} [(s-1)(s-3)p^2 + \alpha_{2s}] \approx \frac{pq}{E(G_s)} \frac{(s-1)(s-3)p^2 + \alpha_{2s}}{\beta_s} \quad (4.82)$$

Les estimations de p selon la méthode de RIDER sont correctes («consistent»); mais elles sont biaisées, du moins pour autant qu'il s'agit de véritables estimations d'après RIDER pour les cas où $k = 1$, $\rho = 1$ et $\rho = 2$. Les estimations d'après RIDER ne sont pas efficaces («efficient») non plus, parce que leurs variances sont plus grandes que dans le cas de la méthode Maximum Likelihood, comme le montre la table IIIa-IIIb. Quoi qu'il en soit, dans le cas où $\rho = 2$, $k = 1$, les variances d'après RIDER sont inférieures et les poids w_s supérieurs, en conséquence, aux grandeurs respectives dans la généralisation de WEINBERG, de sorte que la méthode de RIDER est préférable. La table donne l'efficacité relative J_R de la méthode de RIDER. Ni la méthode de RIDER ni celle de WEINBERG ne sont exhaustives.

LE PROCÉDÉ DE MOORE

Le procédé de MOORE, pas plus que le procédé de RIDER, ne part directement de considérations génétiques, mais du problème de l'estimation du paramètre inconnu p dans une répartition binomiale tronquée, problème qui trouve son application dans la question de l'estimation de la fréquence de caractère dans les fratries en présence d'un matériel soumis à la sélection. C'est pourquoi, dans le procédé de MOORE, il n'est pas non plus question des deux cas limites $k = 0$ et $k = 1$ pour la sélection simple ou complète, mais le procédé se rapportait à l'origine de moins au cas où $k = 1$.

MOORE part de l'identité

$$\sum \frac{r}{s-r+1} \frac{\binom{s}{r} p^r q^{s-r}}{\alpha_s} = \sum \frac{r}{s-r+1} \frac{s(s-1) \dots (s-r+1)}{1 \cdot 2 \dots r} \frac{p^r q^{s-r}}{\alpha_s} \\ = \frac{p}{q} \sum \binom{s}{r-1} \frac{p^{r-1} q^{s-r+1}}{\alpha_s}, \quad (4.83)$$

Lorsque $n_{s,r}$ représente le nombre de fratries de grandeur s avec r sujets atteints, on peut aussi écrire approximativement :

$$\sum \frac{r}{s-r+1} n_{sr} = \frac{p}{q} \sum n_{s,r-1}. \quad (4.84)$$

Dans les deux membres de l'équation, ce sont les fréquences des fratries de même grandeur s qu'on utilise. La relation donne immédiatement l'estimation de p suivante :

$$\hat{p}_s = \sum_{\rho+1}^s \frac{r}{s-r+1} n_{sr} / \left[\sum_{\rho+1}^s \frac{r}{s-r+1} n_{sr} + \sum_{\rho+1}^s n_{s,r-1} \right]. \quad (4.85)$$

Si $\rho = 1$, n_{s0} est inconnu : afin que n_{s0} n'apparaisse pas dans l'estimation $\hat{p}_s$, il importe d'étendre la somme de $r=2$ jusqu'à $r=s$. D'une manière générale, il faut étendre la somme de $r = \rho + 1$ jusqu'à $r = s$, si les ρ classes inférieures de la répartition sont tronquées.

Pour passer aux expressions en les fonctions $f(r)$ et $g(r)$, on considère les contributions de chacune des fratries ; par conséquent, on étendra la somme de 1 à n_s . Pour $\rho = 1$, les contributions des fratries avec $r = 1, 2, 3, \dots, s$ sujets atteints sont :

$$\text{au numérateur } 0, \frac{2}{s-1}, \frac{3}{s-2}, \dots, \frac{s-1}{2}, \frac{s}{1} \\ \text{au dénominateur } 0+1, \frac{2}{s-1}+1, \frac{3}{s-2}+1, \dots, \frac{s-1}{2}+1, \frac{s}{1}+0.$$

Il en va de même pour le cas où $\rho = 2$, de sorte que $f(r)$ et $g(r)$ ont la forme générale,

$$f(r) = \frac{r-r \delta(r-\rho)}{s+1-r}, \quad g(r) = \frac{s+1-r \delta(r-\rho) - \delta(s-r)}{s+1-r}, \quad r \geq \rho \quad (4.86)$$

avec

$$\delta(i) \begin{cases} = 1 & \text{pour } i = 0 \\ = 0 & \text{ailleurs.} \end{cases}$$

En suivant les règles habituelles pour le calcul des espérances, on obtient :

$$\rho = 1, (k=1): E_s(f(r)) = \frac{p}{q} \frac{\alpha_{1s} \cdot p^s}{\alpha_{1s}}, E_s(g(r)) = \frac{1}{q} \frac{\alpha_{1s} \cdot p^s}{\alpha_{1s}} \quad (4.87)$$

$$\text{avec } \alpha_{1s} = 1 \cdot q^s \quad (4.88)$$

$$\rho = 2, (k=1): E_s(f(r)) = \frac{p}{q} \frac{\alpha_{2s} \cdot p^s}{\alpha_{2s}}, E_s(g(r)) = \frac{1}{q} \frac{\alpha_{2s} \cdot p^s}{\alpha_{2s}} \quad (4.89)$$

$$\text{avec } \alpha_{2s} = 1 \cdot q^s - spq^{s-1}$$

Il s'ensuit qu'avec

$$F_s = \sum_{r=1}^{(n_s)} f(r) = \sum_{r=\rho+1}^s \frac{r}{s-r+1} n_{sr} \quad (4.90)$$

$$G_s = \sum_{r=1}^{(n_s)} g(r) = \sum_{r=\rho+1}^s \frac{r}{s-r+1} n_{sr} + \sum n_{s,r-1} \quad (4.91)$$

on a $E(F_s) / E(G_s) = p$, ce qui justifie le procédé de MOORE.

Avant de calculer la variance, il convient d'introduire une désignation particulière pour la seconde des deux portions du dénominateur de $\hat{p}_s$ (voir (4.85)); on écrira donc :

$$\hat{p}_s = F_s / (F_s + H_s) \quad (4.92)$$

$$H_s = \sum_{r=\rho+1}^s n_{s,r-1} = \sum_{r=1}^{(n_s)} h(r) \quad (4.93)$$

avec

$$h(r) = 1 - \delta(s-r), \quad r \geq \rho \quad (4.94)$$

La formule (4.15) deviendra

$$V\left(\frac{F}{F+H}\right) \approx \frac{1}{E^2(F+H)} [q^2 E(F^2) - 2pq E(FH) + p^2 E(H^2)] \quad (4.95)$$

Au biais correspond, selon (4.12)

$$E\left(\frac{F}{F+H}\right) - \frac{E(F)}{E(F+H)} \approx - \frac{1}{E^2(F+H)} [q E(F^2) - (p-q) E(FH) - p E(H^2)] \quad (4.96)$$

Pour calculer $V(\hat{p}_s)$ on cherchera donc, d'après (4.95), tout d'abord

$$E(F_s^2) = \frac{n_s}{\alpha_{\rho s}} \sum_{\rho+1}^s \frac{r^2}{(s-r+1)^2} \binom{s}{r} p^r q^{s-r}.$$

A cet effet, on considérera l'expression du membre droit comme une fonction des deux variables p et q tout en faisant momentanément abstraction de la relation $p+q=1$. Si l'on multiplie $E(F_s^2)$ par q et qu'on dérive partiellement par rapport à q , on obtient :

$$\frac{\partial}{\partial q} \{q E(F_s^2)\} = \frac{n_s}{\alpha_{\rho s}} \sum_{\rho+1}^s \frac{r^2}{s-r+1} \binom{s}{r} p^r q^{s-r}.$$

En multipliant encore une fois par q et en effectuant une nouvelle dérivation, le dénominateur $(s-r+1)$ pourrait également être éliminé; par quelques transformations, on obtient cependant directement :

$$\begin{aligned} \frac{\partial}{\partial q} \{q E(F_s^2)\} &= \frac{n_s}{\alpha_{\rho s}} s \frac{p^2}{q} \sum_{\rho+1}^s \binom{s-1}{r-2} p^{r-2} q^{s-r+1} + \frac{n_s}{\alpha_{\rho s}} \frac{p}{q} \sum_{\rho+1}^s \binom{s}{r-1} p^{r-1} q^{s-r+1} \\ &= \frac{n_s}{\alpha_{\rho s}} s \frac{p^2}{q} \left[(p+q)^{s-1} \cdot p^{s-1} \cdot \sum_{r=2}^0 \binom{s-1}{r-2} p^{r-2} q^{s-r+1} \right] \\ &\quad + \frac{n_s}{\alpha_{\rho s}} \frac{p}{q} \left[(p+q)^s \cdot p^s \cdot \sum_{r=1}^0 \binom{s}{r-1} p^{r-1} q^{s-r+1} \right] \\ &= \frac{n_s}{\alpha_{\rho s}} (s+1) p \left[(p+q)^{s-1} + (p+q)^{s-2} p + \dots + (p+q) p^{s-2} \right. \\ &\quad \left. + p^{s-1} \right] - \frac{n_s}{\alpha_{\rho s}} s p (p+q)^{s-1} - \frac{n_s}{\alpha_{\rho s}} \sum_{r=1}^0 r \binom{s}{r-1} p^r q^{s-r} \\ q E(F_s^2) &= \frac{n_s}{\alpha_{\rho s}} (s+1) p \left[\frac{(p+q)^s}{s} + \frac{(p+q)^{s-1}}{s-1} p + \dots + \frac{(p+q)^2}{2} p^{s-2} \right. \\ &\quad \left. + (p+q)^{s-1} \cdot \frac{p}{s} - \dots - \frac{p^s}{1} \right] - \frac{n_s}{\alpha_{\rho s}} p \left[(p+q)^s \cdot p^s \right] \\ &\quad - \frac{n_s}{\alpha_{\rho s}} p \sum_{r=1}^0 \frac{r}{s-r+1} \binom{s}{r-1} p^{r-1} q^{s-r+1}. \end{aligned}$$

Si l'on divise par q et qu'on utilise $p+q=1$, on aura

$$E(F_s^2) = \frac{n_s}{\alpha_{\rho s}} \frac{p}{q} \left[Y_s - \sum_{r=1}^0 \frac{r}{s-r+1} \binom{s}{r-1} p^{r-1} q^{s-r+1} \right] \quad (4.97)$$

où

$$\gamma_s = (s+1) \left[\frac{1}{s} + \frac{p}{s-1} + \frac{p^2}{s-2} + \dots + \frac{p^{s-1}}{1} - \frac{p^s}{s} - \frac{p^s}{s-1} - \dots - \frac{p^s}{1} \right] - 1 + p^s \quad (4.98)$$

Pour $\rho = 1$, on a

$$E(F_s^2) = \frac{n_s}{\alpha_{1s}} \frac{p}{q} \left[\gamma_s - \frac{q^s}{s} \right] = \frac{n_s}{\alpha_{1s}} \frac{p}{q} \lambda_{1s} \quad (4.99)$$

pour $\rho = 2$

$$E(F_s^2) = \frac{n_s}{\alpha_{2s}} \frac{p}{q} \left[\gamma_s - \frac{q^s}{s} - \frac{2s}{s-1} p q^{s-1} \right] = \frac{n_s}{\alpha_{2s}} \frac{p}{q} \lambda_{2s} \quad (4.100)$$

Il reste encore à calculer $E(F_s H_s)$ et $E(H_s^2)$. On obtient sans difficulté

pour $\rho = 1$

$$E(F_s H_s) = \frac{n_s}{\alpha_{1s}} \frac{p}{q} \left[\alpha_{1s} - s p^{s-1} q - p^s \right] = \frac{n_s}{\alpha_{1s}} \frac{p}{q} \mu_{1s} \quad (4.101)$$

pour $\rho = 2$

$$E(F_s H_s) = \frac{n_s}{\alpha_{2s}} \frac{p}{q} \left[\alpha_{2s} - s p^{s-1} q - p^s \right] = \frac{n_s}{\alpha_{2s}} \frac{p}{q} \mu_{2s} \quad (4.102)$$

De plus,

pour $\rho = 1$

$$E(H_s^2) = \frac{n_s}{\alpha_{1s}} \left[\alpha_{1s} - p^s \right] = \frac{n_s}{\alpha_{1s}} v_{1s}, \quad (4.103)$$

pour $\rho = 2$

$$E(H_s^2) = \frac{n_s}{\alpha_{2s}} \left[\alpha_{2s} - p^s \right] = \frac{n_s}{\alpha_{2s}} v_{2s}. \quad (4.104)$$

Grâce à ces simplifications, on obtient finalement pour $\rho = 1$ ou $\rho = 2$

$$V(\hat{p}_s) = V\left(\frac{F_s}{F_s + H_s}\right) \approx \frac{pq^2 \alpha_{\rho s}}{n_s v_{\rho s}^2} \left[q \lambda_{\rho s} - 2p \mu_{\rho s} + p v_{\rho s} \right] = \frac{pq^2 \alpha_{\rho s} \omega_{\rho s}}{n_s v_{\rho s}^2} \quad (4.105)$$

Pour le biais, on aura de même :

$$E\left(\frac{F_s}{G_s}\right) - p \approx -\frac{pq}{n_s} \frac{\alpha_{\rho s}}{v_{\rho s}^2} \left[q (\lambda_{\rho s} + \mu_{\rho s} - v_{\rho s}) - p \mu_{\rho s} \right]. \quad (4.106)$$

Les espérances et variances ci-dessus se rapportent au cas où $k = 1$. Pour le cas où $k = 0$, il est nécessaire de définir les $f(r)$ et $g(r)$. Tandis que l'on admet la même répartition pour $\rho = 2$, $k = 0$ que pour $\rho = 1$, $k = 1$, en remplaçant toutefois s et r par $s-1$ et $r-1$, on définit pour $\rho = 1$, $k = 0$ les fonctions $f(r)$ et $g(r)$ comme dans la méthode de WEINBERG, quoique l'estimation $\hat{p}_s$ suivant la formule (4.85) soit valable également pour la répartition non tronquée, dont il s'agirait précisément dans le cas des germains du proband lorsque $\rho = 1$, $k = 0$.

Manifestement, l'estimation de MOORE est à nouveau correcte («consistante»), mais cependant biaisée. Elle n'est pas non plus exhaustive («suffisante»). Les variances asymptotiques, pour de grands n_s , dépassent celles de l'estimation du Maximum Likelihood; c'est-à-dire que les valeurs inverses w_s des variances selon MOORE restent inférieures à celles de l'estimation du Maximum Likelihood, ainsi que le montre les tables IVa - IVb. La table VII indique l'efficacité relative de la méthode de MOORE.

COMPARAISON DES DIFFÉRENTS PROCÉDÉS D'ESTIMATION

Une comparaison tant soit peu complète des différents procédés d'estimation devrait tenir compte, en plus de la variance, des biais, si ce n'est même de la répartition des estimations. Dans le présent travail, nous nous limitons cependant à la comparaison essentiellement numérique des variances (asymptotiques) par fratrie (on a posé $n_s = 1$), ou bien de leurs valeurs inverses, c'est-à-dire les poids w_s . Les tables Ia - IVb contiennent les grandeurs w_s pour $p = 0,05$ (0,05) 0,95, $s = 2$ ou 3 (1) 16 pour la méthode du Maximum Likelihood et les procédés de WEINBERG, RIDER et MOORE, et ceci pour $\rho = 1$ et $\rho = 2$. Dans ce cas, on suppose que le degré de sélection k soit toujours égal à 1; le passage à $k = 0$ se fait suivant les règles établies précédemment.

Des quatre procédés étudiés, il n'y a que celui du Maximum Likelihood et les procédés équivalents qui soient (asymptotiquement) efficaces («efficient»). C'est pourquoi il semble naturel de comparer indirectement entre eux les procédés de WEINBERG, RIDER et MOORE non efficaces les trois; on comparera, pour des valeurs correspondantes de s et p , individuellement les variances respectives ou les poids, aux grandeurs analogues obtenues suivant la méthode du Maximum Likelihood. Dans ce but, on calculera le quotient de la valeur w_s pour l'une de ces trois méthodes et de la valeur w_s correspondante de la méthode du Maximum Likelihood, et l'on désignera ce quotient par $J_s(p)$:

$$J_s(p) = \frac{V_s(\text{Maximum Likelihood})}{V_s(\text{Weinberg, Rider ou Moore})}$$

$$= \frac{w_s(\text{Weinberg, Rider ou Moore})}{w_s(\text{Maximum Likelihood})}$$

Ces quotients sont réunis dans les tables V - VII; ils sont tous ≤ 1 . L'allure générale des courbes $J_s(p)$ en fonction de p et de s , et pour les divers procédés d'estimation, est représentée dans les figures.

On s'aperçoit que pour $\rho = 1$ les valeurs w_2 suivant les trois procédés d'estimation de WEINBERG, RIDER et MOORE concordent avec ceux de la méthode du Maximum Likelihood, c'est-à-dire $J_2(p) = \text{const.} = 1$; de même pour $\rho = 2$, $J_3(p) = \text{const.} = 1$ pour tous les trois procédés. Il s'avère que l'efficacité relative $J_s(p)$, tant dans la méthode de WEINBERG que dans celle de RIDER, est relativement faible, particulièrement dans le domaine de $p = 0,25$ important pour les recherches génétiques; ceci est valable notamment pour la généralisation du procédé de WEINBERG à la répartition privée des classes $r = 0$ et $r = 1$ avec $\rho = 2$. Dans la méthode de MOORE au contraire, l'efficacité est assez bonne dans le domaine de $p = 0,25$ et ne devient défavorable que pour de plus grandes valeurs de p . Cette circonstance parle en faveur de l'application du procédé de MOORE à la place de celui de WEINBERG, lorsqu'on ne veut pas renoncer absolument à l'application de méthodes inefficaces, et qu'on ne veut pas utiliser systématiquement les procédés d'estimation du Maximum Likelihood, dès qu'il s'agit, comme c'est habituellement le cas dans les applications génétiques, de fréquences de caractère théoriques comprises entre $p = 0,25$ et $p = 0,50$. Le calcul de l'estimation $\hat{p}_s$ suivant MOORE est un peu plus compliqué, il est vrai; en vue d'établir les variances, on se servira dans les deux cas des tables, afin d'obtenir pour les estimations des erreurs standards utilisables. L'application des formules de variance $pq / (sn_s)$, valables pour des répartitions binomiales non tronquées, serait certainement incorrecte pour des s petits et moyennement grands.

Contrairement à la méthode du Maximum Likelihood ou à la méthode des scores, les procédés d'estimation de WEINBERG, RIDER et MOORE donnent une estimation $\hat{p}_s$ immédiate de la fréquence de caractère p , et ceci uniquement à partir du nombre des sujets atteints dans les fratries de l'échantillon et sans recours a priori à une valeur p de la fréquence de caractère. Les procédés de WEINBERG, RIDER et MOORE s'écartent donc de la méthode des scores en ce qu'une interpolation inverse pour diverses valeurs de p est superflue pour obtenir l'estimation $\hat{p}_s$. Une interpolation pour les grandeurs w_s est par contre nécessaire dans le calcul de la variance pour la valeur $\hat{p}_s$. Soit $w_s(\hat{p}_s)$ le poids par fratrie obtenu de cette façon pour la valeur $\hat{p}_s$; on a $n_s w_s(\hat{p}_s) = W_s(\hat{p}_s)$ pour poids relatif à toutes les n_s fratries et $1/W_s(\hat{p}_s)$ sera le carré de l'erreur standard de l'estimation $\hat{p}_s$.

Ainsi qu'on l'a indiqué déjà, la façon de déduire le biais et la variance de l'estimation $\hat{p}_s$ suivant la méthode du Maximum Likelihood, ou celle de WEINBERG, de RIDER ou de MOORE montre qu'il s'agit d'estimations correctes (*consistant*) quoique en général biaisées. Le biais sera d'autant moins important que n_s sera plus grand. Pour les propriétés des estimations constatées plus haut, de toute façon il s'agit toujours de propriétés asymptotiques; de même, les grandeurs w_s ; c'est-à-dire les inverses des variances sont considérées asymptotiquement. Finalement, répétons une nouvelle fois que des quatre procédés, ce n'est en général que celui du Maximum Likelihood qui fournit des estimations exhaustives.

Pour terminer, étudions encore brièvement les variations de la fonction déterminant l'efficacité relative $J_s(p)$. Comme c'est le cas où $s = 2$, $\rho = 1$ et où $s = 3$, $\rho = 2$, les valeurs des fonctions $f(r)$ et $g(r)$ impliquent que, lorsque $s = 3$, $\rho = 1$ et $s = 4$, $\rho = 2$, les méthodes de RIDER et de MOORE donnent, pour des mêmes valeurs de p et s , aussi des valeurs égales de w_s , et par suite, des valeurs égales de $J_s(p)$. Les fonctions $J_s(p)$ prennent toujours la valeur 1 aux points $p = 0$ et $p = 1$. Le comportement de $J_s(p)$ en ces deux points peut être étudié en écrivant les grandeurs w_s en fonction des puissances de p ou de q seulement, et en faisant tendre p ou q vers zéro dans les quotients correspondants ayant la valeur w_s donnée par le Maximum Likelihood comme dénominateur. On trouve pour le développement de $J_s(p)$ aux environs de $p = 0$ ou de $q = 0$:

WEINBERG:

$$\rho = 1, p = 0: J_s(p) = 1 - \frac{s-2}{3} p, s \geq 2; q = 0: J_s(q) = 1 - \frac{2}{s-1} q, s > 2 \quad (4.107)$$

$$\rho = 2, p = 0: J_s(p) = 1 - (s-3) p, s \geq 3; q = 0: J_s(q) = 1 - \frac{8}{s-1} q, s > 3 \quad (4.108)$$

RIDER

$$\rho = 1, p = 0: J_s(p) = 1 - \frac{s-2}{3} p, s \geq 2; q = 0: J_s(q) = 1 - \frac{2}{s-1} q, s > 2 \quad (4.109)$$

$$\rho = 2, p = 0: J_s(p) = 1 - \frac{s-3}{9} p, s \geq 3; q = 0: J_s(q) = 1 - \frac{2(s-1)}{(s-2)^2} q, s > 3 \quad (4.110)$$

MOORE:

$$\rho = 1, p = 0: J_s(p) = 1 - \frac{(s-5)^2}{12(s-2)} p, s > 2; q = 0: J_s(q) = 1 - \frac{s-1}{2} q, s > 2 \quad (4.111)$$

$$\rho = 2, p = 0: J_s(p) = 1 - \frac{(s-5)^2}{q(s-3)} p, s \geq 3; q = 0: J_s(q) = 1 - \frac{s-1}{2} q, s > 2 \quad (4.112)$$

Dans les méthodes de WEINBERG et de RIDER, la pente de la courbe $J_S(p)$ au point $p = 0$ s'accroît lorsque s croît, tandis qu'au point $p = 1$, la pente tend toujours plus vers la valeur nulle. Cela signifie qu'il y a un minimum $J_S(p)$ toujours plus marqué lorsque s croît, minimum qui devient toujours plus pointu et se déplace vers la gauche, c'est-à-dire vers $p = 0$, tandis que la branche droite de la courbe s'approche toujours plus de $J_S(p) = 1$. En effet, les formules (4.52) - (4.55) ou (4.82) en relation avec (3.10) - (3.13) montrent que, pour une valeur p déterminée, $0 < p < 1$,

$$J_S(p) = \frac{V(\hat{p}_S \text{ selon la méthode du Maximum Likelihood})}{V(\hat{p}_S \text{ selon la méthode de WEINBERG ou de RIDER})}$$

tend, avec s croissant, vers 1.

En ce qui concerne la méthode de MOORE, on constate que la pente, avec s croissant, augmente non seulement au point $p = 0$, mais également au point $p = 1$, et que cette pente croît plus rapidement au point $p = 1$ qu'au point $p = 0$. C'est pourquoi les courbes $J_S(p)$ sont encore relativement plates pour p voisin de zéro et pour s petit, ce qui explique jusqu'à un certain point la bonne efficacité de l'estimation selon MOORE aux environs de la valeur $p = 0,25$. Pour des valeurs de s plus grandes, la courbe $J_S(p)$ décroît relativement rapidement quand $p = 0$, pour croître plus rapidement encore quand $p = 1$; on voit, par la table VII, que $J_S(p)$ pour $s \geq 13$ dans le cas où $\rho = 1$, et pour $s \geq 10$ dans le cas où $\rho = 2$, passe par un maximum relatif lorsque p parcourt le domaine intermédiaire. Ici aussi, l'intérêt est porté vers le comportement de $J_S(p)$ pour $s \rightarrow \infty$ lorsque p est différent de 0 et de 1. Dans la table VII on observe dans un domaine de p de plus en plus étendu un accroissement de la fonction $J_S(p)$ avec s croissant, et l'on peut montrer que dans la méthode de MOORE également, l'efficacité tend vers 1 quand s est croissant et lorsque p est différent de 0 ou de 1. En effet, si l'on désigne par $\psi_S(p)$ l'expression entre accolades de (4.98):

$$\psi_S(p) = \frac{1}{s} + \frac{p}{s-1} + \dots + \frac{p^{s-1}}{1} - \frac{p^s}{s} - \frac{p^s}{s-1} - \dots - \frac{p^s}{1}, \quad (4.113)$$

on reconnaît facilement que $\psi_S(p)$ converge vers une fonction $\psi(p)$ pour $s \rightarrow \infty$, puisque $\psi_S(p)$ est majorée par la série géométrique. En outre, par quelques transformations élémentaires, on obtient la formule de récurrence:

$$\psi_{s+1} = p \psi_s + \frac{1-p^{s+1}}{s+1}.$$

Si l'on exprime ψ_S au moyen de la formule (4.97) par γ_S :

$$\psi_S = \frac{\gamma_S + 1-p^s}{s+1}, \quad (4.114)$$

il s'ensuit tout d'abord

$$\frac{\gamma_{s+1} + 1 \cdot p^s}{s+2} = \frac{p \gamma_s + p \cdot p^{s+1} + 1 \cdot p^{s+1}}{s+1} \quad (4.115)$$

On peut négliger des expressions telles que p^s , sp^s , q^s etc. lorsque s est suffisamment grand, de sorte que γ_s d'après (4.99) peut être remplacé par λ_s , puis μ_s et ν_s d'après (4.101) - (4.102) ou (4.103) - (4.104) par 1. Si l'on désigne alors par ω_s l'expression entre accolades dans (4.105), on a, en négligeant les termes qui tendent vers 0 comme p^s

$$\omega_s = q \gamma_s - p \text{ ou } \gamma_s = \frac{\omega_s + p}{q} \quad (4.116)$$

Si l'on introduit cette expression dans (4.115), on aura après amplification par q

$$\frac{\omega_{s+1} + 1}{s+2} = \frac{p \omega_s + 1}{s+1} \quad (4.117)$$

De là on a

$$(s+1) \omega_{s+1} = p(s+2) \omega_s + 1,$$

c'est-à-dire pour $s \rightarrow \infty$ ω_s est de la forme $\omega_s \sim \frac{1}{sq}$

La variance $V(\hat{p}_s)$ de $\hat{p}_s$ par fratrie, pour $s \rightarrow \infty$ d'après la formule (4.105), à la forme $V(\hat{p}_s) \sim \frac{pq}{s}$.

Ceci correspond, d'après (3.11) ou (3.13), à la variance de l'estimation $\hat{p}_s$ du Maximum Likelihood, de sorte que l'efficacité $J_s(p)$ de la méthode de MOORE est égale à 1 pour s grand, ainsi que nous venons de l'affirmer.

V. L'ESTIMATION DE LA FRÉQUENCE DE CARACTÈRE p À PARTIR DE FRATRIES DE GRANDEUR ARBITRAIRE

Le généticien qui examine le comportement génétique d'un caractère, dispose habituellement de fratries de diverses grandeurs qui peuvent le renseigner sur la fréquence du caractère, et il s'efforcera de parvenir à une estimation de la fréquence de caractère déduite de toutes les fratries. Dans ce chapitre on se propose de discuter diverses possibilités de grouper les fratries de différentes grandeurs, d'illustrer et de comparer ensuite les possibilités à l'aide d'exemples.

MÉTHODES DE WEINBERG, RIDER ET MOORE

Dans le procédé le plus souvent utilisé pour l'estimation de p , celui de WEINBERG, le groupement se faisait jusqu'ici de façon telle qu'on additionnait, sans égard pour la grandeur des fratries, les fonctions $f(r)$ et $g(r)$ [(4.16), (4.21)], qui, lorsqu'on se limitait aux fratries de grandeur s , fournissaient le numérateur F_s et le dénominateur G_s du quotient pour l'estimation de $\hat{p}_s$. C'est ainsi qu'on obtenait l'estimation combinée à partir de la relation

$$\hat{p} = \sum_{(s)} F_s / \sum_{(s)} G_s = F / G \quad (5.1)$$

Ainsi chaque fratrie apportait une contribution $f(r)$ au numérateur et une contribution correspondante $g(r)$ au dénominateur. Dès lors, les estimations $\hat{p}_s$ de p pour des fratries d'une grandeur définie ne présentaient pas d'intérêt. En supposant la même fréquence de caractère p pour chacune des fratries, on déduit immédiatement de (4.17) et de (4.22) que

$$E(F) / E(G) = p. \quad (5.2)$$

La justification du procédé est la même ici que lorsqu'on se limite à des fratries d'une grandeur déterminée. Tandis que le calcul de l'estimation $\hat{p}$ elle-même est simple, le calcul du biais et de la variance de l'estimation, pour laquelle on peut se servir des formules (4.10) et (4.13) de v. MISES, est plus compliqué; car en effet d'après les équations (4.36) - (4.41), chaque grandeur s apporte aux grandeurs $E(F^2)$, $E(FG)$, $E(G^2)$ une contribution différente qui est une fonction relativement compliquée de p et de s ; d'autre part, ces grandeurs $E(F^2)$, $E(FG)$, $E(G^2)$ doivent être calculées pour la valeur $\hat{p}$, donc habituellement pas pour une valeur commode de p . Des constatations équivalentes sont valables pour les procédés de RIDER et de MOORE.

En appliquant ce procédé, le calcul de la variance pour l'estimation combinée ne peut être effectué d'une manière simple que dans le cas où $\rho = 1$, $k = 0$, c'est-à-dire dans le cas de la méthode de réduction. D'après (4.19), l'estimation $\hat{p}_s$ pour une fratrie de grandeur s est la suivante :

$$\hat{p}_s = \frac{r_s - n_s}{n_s (s-1)},$$

où n_s représente le nombre de fratries de grandeur s , r_s le nombre de leurs sujets atteints. Si l'on a des fratries de différentes grandeurs, et qu'on désigne par $N = \sum_{(s)} n_s$ le nombre de toutes les fratries, par $R = \sum_{(s)} r_s$ le nombre de tous les sujets atteints, et par $S = \sum_{(s)} s n_s$ le nombre de tous les enfants dans

les N fratries, en additionnant les numérateurs et les dénominateurs des différents $\hat{p}_s$, on obtiendra l'estimation globale

$$\hat{p} = \frac{R \cdot N}{S \cdot N}. \quad (5.3)$$

D'après les formules (4.36) - (4.38), on calcule

$$E(F^2) = \sum_{(s)} n_s (s-1) p [1 + (s-2) p]$$

$$E(FG) = \sum_{(s)} n_s (s-1)^2 p$$

$$E(G^2) = \sum_{(s)} n_s (s-1)^2$$

d'où l'on tire $E(F^2) - 2pE(FG) + p^2E(G^2) = pq \sum_{(s)} n_s (s-1) = pq(S-N)$,

à cause de $E(G) = \sum_{(s)} n_s (s-1) = S-N$

$$V(\hat{p}) = \frac{pq}{S \cdot N} \quad (5.4)$$

Cette formule, qui est une égalité au sens stricte (et non une approxima-

tion), doit être prise pour la valeur $\hat{p} = (R-N) / (S-N)$, de sorte qu'avec $\hat{q} = 1 - \hat{p} = (S-R) / (S-N)$, on a finalement

$$V(\hat{p}) = \frac{(R-N)(S-R)}{(S-N)^3} \quad (5.5)$$

Le biais de l'estimation $\hat{p}$ dans le cas où $\rho = 1$, $k = 0$ est de nouveau nul.

Au lieu d'additionner, ainsi que nous venons de l'indiquer, le numérateur et le dénominateur de chaque estimation et d'en déduire l'estimation $\hat{p}$, on obtiendra par la suite, dans les procédés de WEINBERG, RIDER et MOORE, l'estimation $\hat{p}$ comme moyenne pondérée des différentes estimations $\hat{p}_s$. En supposant que les $\hat{p}_s$ ne soient que les différentes estimations de la même fréquence de caractère p , des considérations statistiques indiquent l'emploi des valeurs inverses $W_s = n_s / W_s$ des variances des $\hat{p}_s$ comme poids. Si l'on suppose donc que la fréquence de caractère théorique p soit la même pour des fratries de différentes grandeurs, on calculera en principe tous les poids pour la même valeur, à savoir pour la valeur de l'estimation commune $\hat{p}$. Mais comme $\hat{p}$ n'est pas connu d'avance, on fera varier la valeur p , pour laquelle les poids sont calculés, et l'on utilisera comme estimation la valeur $\hat{p}$ pour laquelle la moyenne pondérée des $\hat{p}_s$ coïncide avec $\hat{p}$ elle-même. A l'aide des tables I à IV on peut facilement trouver cette valeur $\hat{p}$, en calculant la moyenne pondérée des $\hat{p}_s$ pour des couples de valeurs se trouvant de part et d'autre de la valeur présumée $\hat{p}$ et en interpolant. Soient p' et p'' deux valeurs telles que $p' < \hat{p} < p''$, on aura les formules

$$\hat{p}' = \frac{\sum_{(s)} W_s(p') \hat{p}_s}{\sum_{(s)} W_s(p')}, \quad \hat{p}'' = \frac{\sum_{(s)} W_s(p'') \hat{p}_s}{\sum_{(s)} W_s(p'')} \quad (5.6)$$

Des deux grandeurs $\hat{p}' - p'$ et $\hat{p}'' - p''$, l'une sera positive, l'autre négative. On peut choisir comme estimation $\hat{p}$ la valeur, calculée par une interpolation linéaire, où la fonction $\hat{p} - p$ s'annule. Bien entendu, on peut également utiliser plusieurs valeurs pour appliquer une interpolation de degré supérieur.

A cause de la relation $V(\hat{p}_s) = 1/W_s$, on aura comme variance de $\hat{p}$ puisque les différentes fratries d'une grandeur définie et à plus forte raison celles de différentes grandeurs sont indépendantes les unes des autres

$$V(\hat{p}) = \frac{\sum_{(s)} W_s^2(\hat{p}) V(\hat{p}_s)}{[\sum_{(s)} W_s(\hat{p})]^2} = \frac{1}{\sum_{(s)} W_s(\hat{p})} = \frac{1}{W(\hat{p})} \quad (5.7)$$

Ce mode uniforme de pondération permet d'obtenir, à partir des différents procédés d'estimation, des estimations de p à peu près concordantes. Dans les exemples que nous allons discuter, le calcul de $\hat{p}$ selon la méthode de WEINBERG conduit à des valeurs à peu près concordantes, aussi bien par l'application du procédé exposé au début, que de celui que nous venons de

décrire. Mais si d'un autre côté dans le procédé de MOORE, on voulait également additionner les numérateurs ou les dénominateurs, on obtiendrait une estimation $\hat{p}$ qui s'écarterait considérablement de l'estimation déterminée à partir de moyennes pondérées. Il va de soi que le procédé, d'après lequel on divise la somme des numérateurs par la somme des dénominateurs peut également être considéré comme la détermination d'une moyenne pondérée des $\hat{p}_s$; les dénominateurs des estimations apparaîtraient alors comme poids. La raison pour laquelle le procédé de WEINBERG, sous cette dernière forme, fournit à peu près le même résultat que l'utilisation des variances inverses comme poids, est que d'une part on peut remplacer les dénominateurs G_s , en première approximation, par leur espérance $E(G_s)$, et que d'autre part, d'après la formule (4.53), du moins pour p petit et s relativement petit, les inverses des variances, soit les W_s , sont plus ou moins proportionnels à $E(G_s)$, c'est-à-dire qu'elles ne dépendent pas beaucoup de p et de s .

MÉTHODE DU MAXIMUM LIKELIHOOD ET MÉTHODE DES SCORES

La continuation logique de la méthode exposée ci-dessus, selon laquelle on calcule la moyenne pondérée des estimations $\hat{p}_s$ pour des fratries de grandeur s , reviendrait ici à résoudre les équations du Maximum Likelihood (3.20) - (3.23), ou bien des équations correspondantes résultant de l'identification de l'observation et de l'espérance mathématique (3.2) - (3.5) pour tous les s ; on formerait ensuite une moyenne pondérée des estimations $\hat{p}_s$ obtenues. Mais les fonctions d'estimation $\hat{p}_s$, calculées uniquement d'après les observations, ne sont pas aussi faciles à obtenir que les estimations correspondantes dans les procédés de WEINBERG, RIDER et MOORE. On arrive toutefois à une unique équation d'estimation pour $\hat{p}$ en identifiant la somme des espérances mathématiques et la somme des observations, et en résolvant l'équation qui en résulte suivant le procédé d'itération par exemple. En ce qui concerne la méthode du Maximum Likelihood à proprement parler, la considération des fratries de différentes grandeurs signifie que P_s est remplacé dans l'équation (3.14) par un produit de telles probabilités (considérées comme fonctions de p) pour les différents s . On remplace L_s (3.15) par

$$\sum_{(s)} L_s = L \text{ et } \frac{\partial L_s}{\partial p} = D_s = 0 \text{ par } \frac{\partial L}{\partial p} = D = 0.$$

Pour des fratries de différentes grandeurs, on additionne donc les contributions comme on le fait pour des fratries de même grandeur, et on parvient à la même équation que celle obtenue par l'identification des sommes des observations et de la somme des espérances mathématiques pour toutes les fratries.

Puisque la connaissance des $b_s(p)$ pour les différentes valeurs de p permet le calcul des scores d_s et par conséquent D_s et D d'après (3.51) pour les

différentes valeurs de p , on peut aussi résoudre l'équation du Maximum Likelihood pour des fratries de différentes grandeurs en calculant la grandeur $D(p)$ pour les différentes valeurs p et en déterminant, par interpolation inverse, la valeur $\hat{p}$ pour laquelle $D(p) = 0$. En raison de l'additivité et de l'indépendance stochastique des D-scores, la variance de la somme des scores pour une valeur p est aussi égale à la somme des variances des scores pour cette valeur p , donc égale à la somme W des poids W_s pour la valeur en question; l'inverse de cette somme pour la valeur $\hat{p}$ donne, à cause de (3.24) la variance $V(\hat{p})$ de l'estimation $\hat{p}$.

Tandis que pour les D-scores, les $\hat{p}_s$ et $\hat{p}$ sont des solutions des équations du Maximum Likelihood, le procédé des y -scores de FINNEY fournit seulement des valeurs approchées pour $\hat{p}_s$ et $\hat{p}$, en donnant, partant d'une approximation préliminaire p de l'équation du Maximum Likelihood une meilleure approximation de $\hat{p}_s$ et de $\hat{p}$ (voir également (3.38) et (3.39)). Bien entendu, ces estimations sont non seulement des fonctions des valeurs r de l'échantillon, mais précisément aussi des fonctions de la valeur initiale p . D'après (3.40), les scores y_s contiennent, par fratrie de grandeur s , les observations r au moyen de d_s sous la forme $r/w_s(p)$. Mais suivant le procédé du Maximum Likelihood, les r doivent pouvoir être additionnés également pour des fratries de grandeurs différentes. C'est ce qu'on obtiendra en considérant les scores pondérés $w_s(p)y_s(p)$ à la place des y_s . Ainsi, tout naturellement, il en résulte une pondération des « estimations » y_s , le poids des y_s étant l'inverse de leur variance, c'est-à-dire w_s , d'après (3.42). De plus, il convient de prendre toutes les variances ou tous les poids, pour la même valeur p pour tous les s . Pour trouver $\hat{p}$, il faudra tout d'abord faire varier p ; à chaque p correspond un score pondéré (pour chaque s):

$$\hat{Y}(p) = \frac{\sum_{(s)} w_s(p) \sum_{(r)} n_{sr} y_{sr}(p)}{\sum_{(s)} n_s w_s(p)} = \frac{\sum_{(s)} W_s Y_s(p)}{\sum_{(s)} W_s(p)} \quad (5.8)$$

et comme estimation $\hat{p}$, on choisit la valeur p pour laquelle $p = \hat{Y}(p)$. La détermination de cette valeur se fait à nouveau en choisissant p' et p'' respectivement de part et d'autre de la valeur présumée $\hat{p}$. Des deux valeurs $\hat{Y}(p') - p'$ et $\hat{Y}(p'') - p''$, l'une sera positive, l'autre négative, et par une interpolation linéaire on trouvera $\hat{p}$. Il faut calculer également la variance $V(\hat{p})$ pour la valeur $\hat{p}$. A cause de l'indépendance stochastique des scores, on a de nouveau

$$V(\hat{p}) = \left(\frac{\sum_{(s)} W_s^2 \cdot \frac{1}{W_s}}{\left(\sum_{(s)} W_s \right)^2} \right)_{\hat{p}} = \left(\frac{1}{\sum_{(s)} W_s} \right)_{\hat{p}} \quad (5.9)$$

Si l'on ne s'intéresse qu'à l'estimation $\hat{p}$ des fratries de grandeur arbi-

traire, on ne formera même pas les scores Y_s et $W_s Y_s$, mais on formera en bloc $[R \cdot N \delta(k)] / pq$ d'après (3.46), pour tous les s à la fois, N étant le nombre total des fratries de toutes grandeurs, et R le nombre des sujets atteints compris dans ces fratries. On'en soustraira

$$\sum_{(s)} n_s b_s(p) = \sum_{(s)} B_s = B, \text{ en prenant les } b_s(p) \text{ dans les tables.}$$

La différence sera divisée par $\sum_{(s)} n_s w_s = \sum_{(s)} W_s = W$, et le quotient représentera $Y(p)$ pour la valeur p .

Dans le cas où $\rho = 1$, $k = 0$, (méthode de réduction), on n'aura pas besoin de faire ces calculs. On se servira de l'estimation (5.3) avec la variance correspondante (5.5) du procédé de WEINBERG. Les poids w_s prennent alors automatiquement leur valeur juste, c'est-à-dire proportionnelle à $(s-1)$.

EXEMPLES D'APPLICATIONS

A l'aide de deux exemples, nous allons discuter la détermination de la fréquence de caractère en tenant compte des diverses hypothèses sur la sélection. Pour expliquer la méthode, nous ferons les calculs pour les deux valeurs du degré de sélection, donc aussi bien pour $k = 0$ que pour $k = 1$, quoiqu'en pratique, l'un des deux modes de sélection sera plus approprié que l'autre.

Le premier exemple se rapporte à un caractère récessif, à savoir la *surdité**. Suivant le mode de sélection, il s'agit d'une «Geschwisterauslese» typique (sélection complète) et on peut poser sans autre $k = 1$. La table 1 a indique les grandeurs n_{sr} pour $s \geq 2$, $r \geq 1$, c'est-à-dire que le nombre minimum de sujets atteints susceptibles d'être observés dans une fratrie est égal à un. Le second exemple se rapporte au *retinoblastome*. Dans ce cas, il s'agit de sujets atteints faisant partie de la première génération de fratries dans laquelle le caractère s'est manifesté, et dont les parents étaient d'apparence normale. Etant donné que des fratries avec un seul cas isolé non déterminé par l'hérédité, peuvent être mélangées à ces fratries dans des proportions inconnues, il convient de ne considérer que les fratries avec au moins deux sujets atteints, de sorte que $\rho = 2$. Puisque ce caractère est très rare, nous avons cherché le matériel dans la littérature, et on peut admettre qu'on peut décrire les conditions de sélection par le degré de sélection $k = 0$ (un proband par fratrie). La table 1 b donne les grandeurs correspondantes n_{sr} pour $s \geq 3$, $r \geq 2$. **

*) L'auteur est redevable des données numériques à M. le professeur E. Hanhart, qui les a publiées en extrait dans le travail «Zur Mendelistischen Auswertung einer 33 Jahre langen Erforschung von Isolaten» paru dans Novant' anni delle leggi mendeliane, Istituto Gregorio Mendel, Roma 1956.

**) Cf. A. KAELIN «Statistische Prüf- und Schätzverfahren für die relative Häufigkeit von Merkmalsträgern in Geschwisterreihen bei einem der Auslese unterworfenen Material, mit Anwendung auf das Retinagliom». Archiv der Julius Klaus-Stiftung 30 (1955) 3/4.

L'estimation de la fréquence de caractère sera tout d'abord expliquée pour les méthodes de WEINBERG, RIDER et MOORE, en ne considérant que des fratries d'une grandeur définie, par exemple $s = 5$. La première chose à faire est de déterminer les fonctions $f(r)$ et $g(r)$ pour le s en question. Lorsqu'on suppose $\rho = 1$, $k = 1$, on a pour $s = 5$ les $f(r)$ et $g(r)$ d'après WEINBERG et RIDER selon (4.21) et (4.68) :

$$\begin{array}{lllll} f(1) = 0, & f(2) = 2, & f(3) = 6, & f(4) = 12, & f(5) = 20, \\ g(1) = 4, & g(2) = 8, & g(3) = 12, & g(4) = 16, & g(5) = 20, \end{array}$$

Pour la surdi-mutité récessive, les n_{sr} correspondants sont, suivant la table 1a

$$n_{51} = 14, \quad n_{52} = 4, \quad n_{53} = 8, \quad n_{54} = 1, \quad n_{55} = 1; \quad n_5 = 28$$

Donc

$$F_5 = 14 \times 0 + 4 \times 2 + 8 \times 6 + 1 \times 12 + 1 \times 20 = 88$$

$$G_5 = 14 \times 4 + 4 \times 8 + 8 \times 12 + 1 \times 16 + 1 \times 20 = 220$$

d'où l'on tire $\hat{p}_5$ suivant la méthode de WEINBERG et RIDER pour la surdi-mutité récessive (tables 2 et 3): $\hat{p}_s = F_5/G_5 = 88/220 = 0.40$. Comme l'estimation globale $\hat{p}$ est voisine de 0.25, on déterminera les poids w_5 pour les valeurs $p = 0.20, 0.25, 0.30$: suivant les tables IIa et IIIa, ces poids par fratrie seront pour $s = 5$:

$$w_5(0.20) = 15.494, \quad w_5(0.25) = 13.985, \quad w_5(0.30) = 13.209.$$

On en tirera les poids $W_s = n_s w_s$ pour les $n_5 = 28$ fratries de grandeur 5

$$W_5(0.20) = 433.821, \quad W_5(0.25) = 391.594, \quad W_5(0.30) = 369.854.$$

(Ces chiffres, comme les autres valeurs de $n_s w_s$ dans les tables 2-6 ont été calculés à partir des valeurs non arrondies de w_s . La différence entre ces résultats et ceux obtenus en utilisant les valeurs w_s des tables, est pratiquement nulle. Le produit

$$W_5(0.20) \hat{p}_5 = 0.40 \times 433.821 = 173.528$$

est la contribution des fratries de grandeur 5 au numérateur de l'estimation globale de la fréquence de caractère pour la valeur $p' = 0.20$, le poids $W_5(0.20) = 433.821$ est la contribution correspondante au dénominateur. Si l'on divise la somme des contributions du numérateur par la somme de celles du dénominateur pour la valeur $p' = 0.20$, on a pour cette valeur p' l'estimation $\hat{p}' = 737.280/3054.686 = 0.24136$. De même, on aurait pour la valeur $p'' = 0.25$: $\hat{p}'' = 669.423/2777.492 = 0.24102$. D'où l'on déduit que $\hat{p}' - p' = 0.04136$ pour la valeur $p' = 0.20$ et $\hat{p}'' - p'' = -0.00898$ pour la valeur $p'' = 0.25$, donc par interpolation linéaire

$$\hat{p} = (898 \times 0.20 + 4136 \times 0.25) / (898 + 4136) = 0.24108.$$

Pour cette même valeur $\hat{p}$, on calculera également le poids $W(\hat{p})$, dont on déduira l'erreur standard $s_{\hat{p}}$ de l'estimation $\hat{p}$ sous la forme $1/\sqrt{W(\hat{p})}$. On obtient

$$W(\hat{p}) = [(0.25 - 0.24108) \times 3054.686 + (0.24108 - 0.20) \times 2777.492] / 0.05 \\ = 2826.943$$

De là on a $s_{\hat{p}} = 1/\sqrt{2826.943} = 0.01881$, de sorte que l'estimation de la fréquence de caractère pour la surdi-mutité récessive sera, en admettant une « Geschwisterauslese » selon WEINBERG et RIDER :

$$\hat{p} = 0.24108 \pm 0.01881$$

Lorsqu'il s'agit, en premier lieu, de vérifier l'hypothèse $p = 0.25$, on utilisera le poids $W = 2777.492$ pour la valeur $p = 0.25$ pour calculer l'erreur standard $s_{\hat{p}}$: $s_{\hat{p}} = 0.01897$. Si l'on néglige une éventuelle dissymétrie de la répartition de l'estimation $\hat{p}$, on peut vérifier l'hypothèse $p = 0.25$, en calculant l'écart $u = |0.24108 - 0.25| / 0.01897 = 0.43$, ce qui n'est naturellement pas significatif.

De même que dans la méthode de WEINBERG ou celle de RIDER, on peut déterminer les estimations d'après MOORE. Il s'agit ici aussi d'utiliser les fonctions appropriées $f(r)$ et $g(r)$ afin d'en déduire F_s et G_s , puis de là de nouveau $\hat{p}_s$. On peut tirer des tables les poids w_s par fratrie. Dans le cas du rétinoblastome où l'on a affaire à une répartition privée des classes $r = 0$ et $r = 1$, en supposant une sélection complète ($k = 1$), pour $s = 5$ d'après (4.86), les fonctions f et g prendront la valeur :

$$f(2) = 0, \quad f(3) = \frac{3}{3}, \quad f(4) = \frac{4}{2}, \quad f(5) = \frac{5}{1} \\ g(2) = 1, \quad g(3) = \frac{6}{3}, \quad g(4) = \frac{6}{2}, \quad g(5) = \frac{5}{1}$$

A l'exception du cas où $r = s$, où f et g prennent la même valeur, c'est-à-dire s , g dépasse toujours de 1 le f correspondant. Dans la table 1b se trouvent les grandeurs correspondantes n_{sr}

$$n_{52} = 1, \quad n_{53} = 2, \quad n_{54} = 2, \quad n_{55} = 0 \quad . \quad n_5 = 5$$

Cela donne $F_5 = 6$, $G_5 = 11$, donc $\hat{p}_5 = 6/11 = 0.54545$. Les valeurs des poids w_5 par fratrie sont, lorsque $p' = 0.30$ et $p'' = 0.35$ d'après la table IVb $w_5(0.30) = 8.431$, ou $w_5(0.35) = 8.459$, de sorte que les contributions au dénominateur de l'estimation globale sont 42.157 ou 42.295, tandis que les contributions aux numérateurs se montent à $0.54545 \times 42.157 = 22.995$ ou à $0.54545 \times 42.295 = 23.070$. Les autres calculs sont semblables à ceux utilisés dans le procédé de WEINBERG; quant aux détails, on peut les trouver dans la table 4.

Les tables 2-4 contiennent, en plus, les totaux $\sum_{(s)} F_s$ et $\sum_{(s)} G_s$ des numé-

rateurs et dénominateurs des estimations individuelles et, entre parenthèses, l'estimation qui en a été déduite $\bar{p} = \sum_{(s)} F_s / \sum_{(s)} G_s$. Selon le procédé d'estimation, on a des déviations par excès, en partie assez considérables, des estimations décrites précédemment. La raison en est que la répartition des « poids » s'écarte de celle des poids W_s . Indépendamment de ces écarts, le calcul de la variance de $\bar{p}$ présenterait certaines difficultés.

Il faut encore brièvement expliquer le cas de la sélection simple ($k = 0$). Les choses se présentent simplement lorsque $\rho = 1$. Pour les méthodes de RIDER et de MOORE, et, après un passage à la limite, pour la méthode de WEINBERG, on a par définition

$$f(r) = r-1, \quad g(r) = s-1 \quad (\text{cf. (4.19)}),$$

donc

$$\hat{p}_s = \frac{r_s - n_s}{n_s (s-1)},$$

d'autre part, on a, d'après (4.52) w_s par fratrie $= (s-1)/pq$, de sorte que, si $N = \sum_{(s)} n_s$, $R = \sum_{(s)} m_s$, $S = \sum_{(s)} sn_s$, l'estimation globale $\hat{p}$ et le poids

correspondant sont :

$$\hat{p} = \frac{R-N}{S-N}, \quad W(p) = \frac{S-N}{pq}, \quad W(\hat{p}) = \frac{(S-N)^3}{(R-N)(S-R)}.$$

Pour la surdi-mutité récessive, on trouve dans la table 1 $N = 150$, $R = 267$, $S = 910$, donc

$$\hat{p} = \frac{117}{760} = 0.15395, \quad V(\hat{p}) = \frac{117 \times 643}{(760)^3} = 1.71381 \times 10^{-4},$$

finalement

$$\hat{p} = 0.15395 \pm 0.01309.$$

L'estimation $\hat{p} = 0.154$ est très petite, et l'hypothèse d'une sélection simple (par proband) n'était en effet pas justifiée dans ce cas.

Conformément à une convention antérieure, si $k = 0$, mais $\rho = 2$, on utilisera le procédé correspondant en calculant cependant partout (excepté naturellement pour les n_{sr}) les fonctions f et g et les poids w pour $r-1$ et $s-1$ à la place de r et s . Dans le cas du rétinoblastome, lorsque $s = 5$, on doit prendre les valeurs f et g pour $s = 4$ et remplacer chaque fois r par $r-1$. On trouve ainsi d'après la méthode de MOORE

$$f(2) = 0, \quad f(3) = \frac{2}{3}, \quad f(4) = \frac{3}{2}, \quad f(5) = \frac{4}{1}$$

$$g(2) = 1, \quad g(3) = \frac{5}{3}, \quad g(4) = \frac{5}{2}, \quad g(5) = \frac{4}{1}$$

$$n_{52} = 1, \quad n_{53} = 2, \quad n_{54} = 2, \quad n_{55} = 0, \quad n = 5,$$

donc $F_5 = 4.33$, $G_5 = 9.33$, $\hat{p}_5 = 0.46429$. Comme poids, on utilise les valeurs obtenues pour $s = 4$ ($\rho = 1$, $k = 1$), c'est-à-dire pour la valeur $p' = 0.25$: $w_5(0.25) = 10.863$, pour la valeur $p'' = 0.30$: $w_5(0.30) = 10.091$, d'où $W_5(0.25) = 54.316$, $W_5(0.30) = 50.455$. Les autres calculs se poursuivent comme précédemment.

Nous examinerons deux possibilités de grouper les fratries dans la méthode du Maximum Likelihood. La première consiste à déterminer, pour chaque s , l'estimation $\hat{p}_s$, en résolvant l'équation correspondante du Likelihood (cf. (3.2) - (3.5) ou (3.20) - (3.23)). Comme dans la méthode de WEINBERG, RIDER ou MOORE, $\hat{p}_s$ ne sera fonction que des valeurs r de l'échantillon dans des fratries de grandeur s , et les autres calculs peuvent être effectués exactement sur le modèle des méthodes dont nous venons de parler. Les détails du calcul se trouvent dans la table 5. La résolution des équations du Likelihood est, en général, fastidieuse, et le plus souvent, ce sont moins les estimations individuelles $\hat{p}_s$ qui nous intéressent que l'estimation globale $\hat{p}$ pour des fratries de n'importe quelle grandeur. Pour obtenir cette dernière, on se sert de la méthode des Maximum Likelihood scores de FINNEY. Partant d'une valeur commode p , comme par exemple $p = 0.25$, et en se limitant tout d'abord à des fratries d'une grandeur déterminée telle que $s = 5$, on trouvera une solution approximative Y_s de l'équation du Likelihood pour des fratries de cette grandeur. Du point de vue purement arithmétique, ce ne sont pas ces grandeurs Y_s qu'on obtient directement, mais c'est le score pondéré $W_s Y_s$ pour des fratries de grandeur s qui est intéressant. Le score pondéré $W_s Y_s$ est la somme $\sum w_s y_s$ des scores pondérés pour les fratries individuelles de grandeur s pour une valeur déterminée p , conformément à (3.46), de sorte que

$$W_s Y_s = \sum w_s y_s = \frac{r_s - n_s \delta(k)}{p q} - n_s b_s(p) \quad (5.10)$$

En additionnant les valeurs r de chaque fratrie, on obtient les valeurs r_s , et les grandeurs $b_s(p)$ et $w_s(p)$ sont contenues dans les tables Ia - Id pour des valeurs appropriées de p et n'ont qu'à être multipliées par n_s . Ainsi on a pour $s = 4$, $p = 0.20$ ($\rho = 1$, $k = 1$) : $w_4(0.20) = 12.967$, $b_4(0.20) = 5.875$. Pour les $n_4 = 20$ fratries avec $r_4 = 30$ sourds-muets (table 1a), on a donc pour la valeur $p = 0.20$

$$W_4 = n_4 w_4 = 259.344, \quad B_4 = n_4 b_4 = 117.500,$$

donc
$$W_4 Y_4(0.20) = \frac{30}{0.16} - 117.50 = 187.5 - 117.5 = 70.0 = 259.344 Y_4$$

En plus de l'approximation initiale $p = 0.20$ de la solution de l'équation du Maximum Likelihood, on obtient par la même occasion l'approximation suivante :

$$Y_4 = 70.0 / 259.344 = 0.26991.$$

Pour obtenir une estimation globale $\hat{p}$ pour des fratries de grandeur arbitraire réunies, on a besoin, suivant (5.8), de $\sum_{(s)} W_s Y_s$ et de $\sum_{(s)} W_s$ pour certaines valeurs de p . Pour ce faire, il n'est pas nécessaire de noter séparément, comme dans la table 6, les $W_s Y_s$ et W_s pour les différents s . Afin d'obtenir $W(p) = \sum_{(s)} W_s(p)$, il suffit d'additionner les produits n_s par w_s pour la valeur p (table Ia - Ib). Pour obtenir $\sum_{(s)} W_s Y_s$, on forme d'après (5.10) la somme $B(p)$ des produits n_s par $b_s(p)$, suivant la table Ic - Id; si on la soustrait de $[R - N \delta(k)] / pq$, on a $WY = \sum_{(s)} W_s Y_s$ pour la valeur p , et on peut en déduire l'estimation $\hat{Y}$. Pour la surdi-mutité récessive, on a pour une valeur initiale $p' = 0.20$, et lorsque la sélection est complète ($k = 1$, $\rho = 1$)

$W = 3811.975$, $R = 267$, $R/pq = 267/0.1600 = 1668.750$, $B = 795.334$
donc

$$p' = 0.20 : W' \hat{Y}' = 1668.750 - 795.334 = 873.416, \quad W' = 3811.975 \\ \hat{Y}' = 0.22912$$

De façon analogue

$$p'' = 0.25 : W'' \hat{Y}'' = 1424.000 - 619.071 = 804.929, \quad W'' = 3518.378 \\ Y'' = 0.22878$$

En vue de l'interpolation, on calcule les valeurs analogues à $\hat{p} - p \hat{Y} - p$. Pour $p = p' = 0.20$, on a $\hat{Y}' - p' = 0.02912$, de même on a $\hat{Y}'' - p'' = -0.02122$ pour la valeur $p'' = 0.25$. En interpolant linéairement, on trouve

$$\hat{p} = (0.02122 \times 0.20 + 0.02912 \times 0.25) / (0.02122 + 0.02912) = 0.22892$$

Si l'on cherche la variance de $\hat{p}$, on calcule, par interpolation linéaire, le poids W pour la valeur $\hat{p} = 0.22892$. On a $W(\hat{p}) = 3642.158$ et de là $s_{\hat{p}} = 1/\sqrt{W(\hat{p})} = 0.01657$.

Pour compléter, on peut encore indiquer les grandeurs B et W pour le rétinoblastome ($\rho = 2$) dans le cas des sélections complète et simple ($k = 1$ ou $k = 0$):

$$\rho = 2, k = 1 : p' = 0.35 : B' = 262.636, \quad W' = 615.763$$

$$p'' = 0.40 : B'' = 227.812, \quad W'' = 640.114$$

$$\rho = 2, k = 0 : p' = 0.30 : B' = 134.309, \quad W' = 710.312$$

$$p'' = 0.35 : B'' = 111.646, \quad W'' = 693.657$$

On trouve les n_s et les grandeurs N , R et $R - N$ dans la table 1.

Il n'y a pas lieu d'examiner le cas où $\rho = 1$, $k = 0$ (sélection simple) en ce qui concerne la surdi-mutité récessive, puisqu'il a été traité lors de la discussion du procédé de WEINBERG, qui coïncide dans ce cas avec la méthode du Maximum Likelihood.

La table 7 donne une classification des estimations (interpolation linéaire) de même que leur erreur standard. Les erreurs standard pour la valeur $p = 0.25$ permettent de comparer les variances pour les différents procédés et de vérifier en même temps l'hypothèse $p = 0.25$. Dans le cas de la surditité, les écarts entre l'observation et l'hypothèse d'une hérédité récessive restent fortuits lorsque $k = 1$. L'hypothèse $k = 0$ avec une estimation de $\hat{p} = 0.154$ est en contradiction avec les conditions de sélection, et les résultats n'ont été indiqués qu'à titre d'illustration numérique complète. Dans le rétinoblastome, les écarts entre l'hypothèse $p = 0.25$ et l'observation sont également fortuits, si l'on suppose la sélection simple. Avec l'hypothèse de la sélection complète ($k = 1$), pourtant peu fondée lorsqu'on considère le matériel d'observation collectionné dans la littérature, on obtiendrait, en appliquant la méthode de MOORE et celle du Maximum Likelihood, un écart significatif, en suivant les méthodes de WEINBERG et de RIDER, un écart fortuit à partir de 0.25.

RÉSUMÉ

Le prélèvement des fratries se pratique couramment par les sujets atteints qu'elles contiennent, en particulier pour des caractères récessifs rares; il entraîne certains effets de sélection dont on peut tenir compte en utilisant des répartitions binomiales privées de certaines classes au lieu de répartitions binomiales complètes. La façon de tronquer dépend du mode de sélection (sélection simple ou complète; $k = 0$ ou $k = 1$; nombre des probands égal à 1 ou bien égal au nombre de sujets atteints dans la fratrie sélectionnée) et du nombre minimum ρ ($= 1$ ou $= 2$) de sujets atteints nécessaire pour qu'une fratrie puisse être prélevée.

On cherche une estimation de la fréquence d'un caractère dans des fratries, estimation corrigée quant aux effets de sélection. Dans la littérature médico-génétique, le procédé d'estimation de WEINBERG est le plus connu, tandis que les procédés d'estimation plus récents de RIDER et de MOORE et, en particulier, celui du Maximum Likelihood de FISHER et FINNEY, en dépit de certains avantages, n'ont connu jusqu'à présent qu'une diffusion limitée. Le présent travail apporte la justification et une comparaison des divers procédés, en insistant sur la détermination et la comparaison des variances des estimations, tout en indiquant le calcul de l'espérance mathématique des estimations et celui des biais qui peuvent s'ensuivre.

Exception faite du cas de la sélection simple ($k = 0$) avec un minimum de 1 sujet atteint ($\rho = 1$), dans lequel cas toutes les quatre méthodes donnent des résultats identiques (méthode de réduction de JUST), ni le procédé de

WEINBERG, ni ceux de RIDER et MOORE ne sont efficaces («efficient») du point de vue statistique: les estimations de la fréquence de caractère dans des fratries d'une grandeur définie s donnent des variances plus grandes que les estimations correspondantes suivant la méthode du Maximum Likelihood. Le calcul des efficacités pour des fréquences de caractère théorique déterminées ($p = 0.05$ (0.05) 0.95), et pour des grandeurs s de fratries déterminées ($s = 2$ ou 3 (1) 16) fait ressortir une certaine supériorité de la méthode de MOORE par rapport à celles de WEINBERG et RIDER, dans la répartition privée de la classe $r = 0$ ($\rho = 1$) et à plus forte raison dans la répartition privée des classes $r = 0$ et $r = 1$ ($\rho = 2$) tout particulièrement dans le domaine, important du point de vue génétique, aux environs de la valeur $p = 0.25$.

Les estimations suivant les quatre procédés mentionnés sont biaisées, exception faite du cas ($k = 0$, $\rho = 1$). Par contre, les estimations sont correctes («consistent»). De plus, les estimations selon le Maximum Likelihood sont exhaustives («sufficient»), tandis que celles faites d'après les procédés de WEINBERG, RIDER et MOORE ne le sont pas, exception faite du cas mentionné.

Du point de vue ces calculs numériques, les procédés de WEINBERG, RIDER ou MOORE pourraient éventuellement paraître plus simples que le procédé du Maximum Likelihood, parce que, partant directement des observations, ils fournissent une estimation immédiate de la fréquence de caractère. Cet avantage tombe lorsqu'il s'agit de calculer l'erreur standard appropriée correspondant à l'estimation. Le calcul des erreurs standard des estimations d'après WEINBERG, RIDER ou MOORE est facilité par des tables des valeurs inverses $w_s(p)$ des variances pour des valeurs appropriées de s et p , et pour les répartitions privées soit d'une classe, soit de deux ($\rho = 1$ ou $\rho = 2$). Par interpolation, on peut trouver les variances pour les valeurs intermédiaires. La présentation sous forme de tables des valeurs inverses w_s des variances pour des couples de valeurs (s ; p) est conditionnée par le mode de formation d'une estimation combinée de la fréquence de caractère à partir des fratries de différentes grandeurs. Nous écartant du procédé habituel de WEINBERG par exemple, nous proposons comme estimation combinée de fréquence de caractère une moyenne pondérée des estimations pour chacune des grandeurs s des fratries, en choisissant comme poids les inverses des variances. Il convient de prendre les poids toujours pour la même valeur p , qu'on fera varier jusqu'à ce que la moyenne pondérée coïncide avec la valeur initiale p . A l'aide des tables, on peut trouver cette valeur par interpolation linéaire avec une précision suffisante et avec un minimum de calculs. La variance de l'estimation combinée $\hat{p}$ de la fréquence de caractère est l'inverse de la somme des poids pour la valeur $\hat{p}$. Cette façon de grouper des fratries de différentes grandeurs devrait conduire à des valeurs à peu près concordantes des estimations combinées.

Le procédé indiqué pour grouper les fratries de différentes grandeurs est analogue à la détermination d'une estimation commune au moyen des scores suivant la méthode du Maximum Likelihood. Cependant, au lieu d'utiliser pour toutes les grandeurs de fratries l'estimation de la fréquence de caractère obtenue en résolvant l'équation du Likelihood correspondante, on emploie, pour tous les s une première approximation («Maximum Likelihood Score») de la solution de l'équation du Likelihood correspondante (obtenue d'après le procédé de NEWTON), en partant d'une valeur initiale commune p choisie convenablement. Comme estimation combinée de la fréquence de caractère pour des fratries de toutes les grandeurs, on utilise la valeur p pour laquelle le score pondérée coïncide avec la valeur initiale, les poids étant pris à la valeur p . A cause du caractère additif des scores (pondérés) et de la propriété des estimations du Maximum Likelihood d'être exhaustives, de sorte qu'à part le nombre de fratries de différentes grandeurs il n'y a que le nombre total des sujets atteints dans toutes les fratries qui entre dans les calculs, la méthode des scores est particulièrement simple du point de vue des calculs lorsqu'on utilise des tables appropriées.

Sans examiner la portée de ces calculs du point de vue de la génétique humaine, on a comparé les différents procédés d'estimation en les illustrant à l'aide de l'exemple de la surdi-mutité récessive (pour la répartition binomiale privée d'une classe, $\rho = 1$), ou de l'exemple du rétinoblastome dans des fratries avec parents d'apparence normale (répartition binomiale privée de deux classes, $\rho = 2$). On a indiqué les estimations avec les erreurs standard en partant de l'hypothèse d'une sélection simple ($k = 0$), ou d'une sélection complète ($k = 1$).

SUMMARY

Genetic sibships are often traced by starting from affected individuals, especially when the investigation concerns rare recessive traits. This entails selection effects which can be allowed for by employing the truncated binomial distribution. The extent of truncation depends on the method of selection simple or complete (i.e. the number of probands being equal to 1 ($k = 0$); or to the number of affected individuals in a selected sibship ($k = 1$)) and on the smallest number ρ (1 or 2) of affected individuals deemed necessary in order that the sibship be included in the sample.

We require an estimate of the frequency of a trait in sibships, corrected for these effects of selection. In medico-genetical literature WEINBERG's method of estimation is the best known, whereas the more recent methods of RIDER and of MOORE, and particularly the Maximum Likelihood method of FISHER and its adaptation by FINNEY have not, in spite of certain advantages, been extensively used. The present paper deals with these methods of estimation, comparing their standard errors and biases, and gives computational details.

Apart from the case of simple selection ($k = 0$) with at least one affected individual ($\rho = 1$), where all four methods yield identical results (JUST's «Reduktionsmethode») neither WEINBERG's method nor those of RIDER and of MOORE are efficient from the statistical viewpoint; these frequency estimates for sibships of a given size s will give larger variances than the corresponding estimates by the Maximum Likelihood method. The calculation of the efficiencies for given theoretical trait frequencies ($p = 0.05$ (0.05) 0.95) and for given sizes s of sibship ($s = 2$, or 3 (1) 16) shows that MOORE's method is superior to the methods of WEINBERG and of RIDER in the genetically important range of values around $p = 0.25$, when $\rho = 1$ and particularly so when $\rho = 2$.

The estimates by the four above mentioned methods are all biased, apart from the case $k = 0$, $\rho = 1$. On the other hand, these estimates are consistent. Moreover, the Maximum Likelihood estimates are sufficient, while those made by the methods of WEINBERG, of RIDER and of MOORE are not, except in the case $k = 0$, $\rho = 1$.

As far as computation is concerned, the methods of WEINBERG, of RIDER and of MOORE might seem simpler than the Maximum Likelihood method as they provide a direct estimate of the trait frequency based only on the observations. This advantage carries less weight when the appropriate standard error of the estimate is to be calculated. The calculation of standard errors for the estimates of WEINBERG, RIDER or MOORE is made easier by the tabulation of the inverse values $w_s(p)$ of the variances for appropriate values of s , p and ρ . The choice of the values w_s for tabulation, rather than of the variances themselves is due to the manner in which sibships of different sizes are combined to form an overall estimate of the trait frequency. As against the method of WEINBERG, for instance, we propose a combined estimate of the trait frequency as a weighted average of the estimates for each sibship size s , choosing for weights w_s . These weights have to be taken always for the same value p which is varied until the weighted mean coincides with the initial value p . With the aid of the tables given, this value can be found with sufficient accuracy and with a minimum of computation by linear interpolation. The variance of the combined estimate of the trait frequency is the inverse of the summed weights corresponding to the final estimate of p . This way of grouping sibships of different sizes should lead to almost the same value of the combined estimates.

The method advocated for grouping sibships of various sizes is analogous to estimation by means of Maximum Likelihood scores. However, instead of employing for each size of sibship, the frequency estimate obtained by solving the corresponding Maximum Likelihood estimate equation, a first approximate solution (Maximum Likelihood score) is obtained for each s , starting from a common and convenient initial value p . The value $\hat{p}$ for which the

weighted score of the estimates coincides with the initial value, the weights being taken at $\hat{p}$, is used as the combined estimate for sibships of all sizes. The (weighted) scores being additive and the Maximum Likelihood estimates being sufficient, only the total number of affected individuals in all sibships enters into the calculations, apart from the number of sibships of different sizes. Thus, the method of scores is especially simple from the computational viewpoint, with the tables given in this paper.

The methods are illustrated by calculations based on data concerning recessive deaf-mutism and retinoblastoma. The estimates with their standard errors are computed for the hypotheses of both simple and complete selection.

ZUSAMMENFASSUNG

Die Erfassung von Geschwisterreihen durch die in ihnen enthaltenen Merkmalsträger, wie sie besonders bei seltenen rezessiven Merkmalen üblich ist, hat Auslesewirkungen zur Folge, denen durch die Verwendung abgeschnittener Binomialverteilungen, in denen gewisse Klassen fehlen, an Stelle vollständiger Binomialverteilungen Rechnung getragen werden kann. Die Art der Abschneidung hängt ab von der Art der Erfassung (einfache oder vollständige Auslese; $k = 0$ bzw. $k = 1$; Anzahl der Probanden gleich Eins oder gleich der Zahl der Merkmalsträger in der erfassten Geschwisterreihe) und von der Minimalzahl ρ ($= 1$ oder $= 2$) der Merkmalsträger, die nötig ist, damit eine Geschwisterreihe überhaupt erfasst wird.

Gesucht ist eine bezüglich der Auslesewirkungen korrigierte Schätzung der Häufigkeit eines Merkmals in Geschwisterreihen. In der medizinisch-genetischen Literatur am bekanntesten ist das Schätzverfahren von WEINBERG, während die neuern Schätzverfahren von RIDER und MOORE und insbesondere dasjenige der Maximum Likelihood nach FISHER und FINNEY trotz gewisser Vorzüge bisher geringe Verbreitung gefunden haben. Die vorstehende Arbeit bringt die Begründung und einen Vergleich der verschiedenen Verfahren, wobei neben der Berechnung des Erwartungswertes der Schätzungen und im Zusammenhang damit allfälliger einseitiger Fehler besonders auf die Berechnung und den Vergleich der Streuungen der Schätzungen Gewicht gelegt wird.

Abgesehen vom Fall der einfachen Auslese ($k = 0$) mit einer zur Erfassung einer Geschwisterreihe nötigen Mindestzahl von 1 Merkmalsträger ($\rho = 1$), wo alle vier Methoden zu gleichlautenden Ergebnissen führen (JUSTsche Reduktionsmethode), sind weder das Verfahren von WEINBERG noch diejenigen von RIDER und MOORE im statistischen Sinne wirksam («efficient»), d.h. die Schätzungen der Merkmalshäufigkeit in Geschwisterreihen einer bestimmten Grösse s nach diesen Verfahren sind mit grösseren Streuungen behaftet als die entsprechenden Schätzungen nach Maximum Likelihood. Die Berech-

nung der Wirkungsgrade für bestimmte theoretische Merkmalshäufigkeiten ($p = 0.05$ (0.05 0.95) und für bestimmte Grössen s von Geschwisterreihen ($s = 2$ bzw. 3 (1 16)) zeigt für die einfach ($\rho = 1$), besonders aber für die doppelt ($\rho = 2$) abgeschnittene Binomialverteilung gerade im genetisch wichtigen Bereich $p = 0.25$ eine gewisse Ueberlegenheit der Methode von MOORE über diejenigen von WEINBERG und RIDER.

Die Schätzungen nach allen vier erwähnten Schätzverfahren sind, vom erwähnten Ausnahmefall ($k = 0$, $\rho = 1$) abgesehen, nicht erwartungstreu, d. h. sie sind mit einem einseitigen Fehler behaftet. Hingegen sind die Schätzungen passend («consistent»). Ferner erweisen sich die Maximum Likelihood Schätzungen, nicht jedoch, bis auf die erwähnte Ausnahme, diejenigen nach den Verfahren von WEINBERG, RIDER oder MOORE als erschöpfend («sufficient»).

In rechnerischer Beziehung mögen die Verfahren von WEINBERG, RIDER oder MOORE vielleicht einfacher erscheinen als das Maximum Likelihood-Verfahren, weil sie unmittelbar Schätzungen der Merkmalshäufigkeit direkt auf Grund der Beobachtungen liefern. Dieser Vorzug fällt dahin, wenn es sich um die Berechnung des richtigen zur Schätzung gehörigen Standardfehlers handelt. Die Berechnung der Standardfehler der Schätzungen nach WEINBERG, RIDER oder MOORE wird erleichtert durch Tabellen mit den reziproken Werten $w_s(p)$ der Streuungen für passende Werte von s und p und für die einfach und doppelt abgeschnittenen Verteilungen ($\rho = 1$ bzw. $= 2$). Streuungen für Zwischenwerte können durch Interpolation erhalten werden. Die Tabellierung der reziproken Werte w_s der Streuungen für passende Wertepaare (s ; p) ist begründet in der Art der Bildung einer kombinierten Schätzung der Merkmalshäufigkeit auf Grund von Geschwisterreihen verschiedener Grösse. Abweichend vom üblichen Vorgehen von WEINBERG zum Beispiel wird als kombinierte Schätzung der Merkmalshäufigkeit ein gewogenes Mittel der Schätzungen für die einzelnen Grössen s der Geschwisterreihen vorgeschlagen, wobei als Gewichte die reziproken Werte der Streuungen, d. h. die Grössen w_s gewählt werden. Die Gewichte sind einheitlich an derselben Stelle p zu nehmen, die so lange variiert wird, bis das gewogene Mittel mit dem Ausgangswert p zusammenfällt. Mit Hilfe der Tabellen kann dieser Wert mit genügender Genauigkeit und mit geringem Rechenaufwand durch lineare Interpolation gefunden werden. Die Streuung der kombinierten Schätzung $\hat{p}$ der Merkmalshäufigkeit ergibt sich als reziproker Wert der Summe der Gewichte an der Stelle $\hat{p}$. Diese Art der Zusammenfassung von Geschwisterreihen verschiedener Grösse dürfte zu einigermaßen übereinstimmenden Werten der kombinierten Schätzungen führen.

Das angedeutete Verfahren der Zusammenfassung von Geschwisterreihen verschiedener Grösse steht in Analogie zur Ermittlung einer gemeinsamen Schätzung nach der Methode der Maximum Likelihood Scores. Statt jedoch für

jede Grösse von Geschwisterreihen die durch Auflösung der zugehörigen Likelihood-Gleichung erhaltene Schätzung der Merkmalshäufigkeit zu verwenden, wird, ausgehend von einem bequemen gemeinsamen Ausgangswert p , zu jedem s (nach dem NEWTONschen Verfahren) eine erste Näherung (Maximum Likelihood Score) der Lösung der entsprechenden Likelihood-Gleichung benützt. Als kombinierte Schätzung der Merkmalshäufigkeit für Geschwisterreihen aller Grössen wird dasjenige p verwendet, für welches der gewogene Score, die Gewichte an der Stelle p genommen, mit dem Ausgangswert zusammenfällt. Wegen der Additivität der (gewogenen) Scores und wegen der Eigenschaft der Maximum Likelihood Schätzungen, erschöpfend zu sein, so dass neben den Anzahlen von Geschwisterreihen verschiedener Grösse nur die Gesamtzahl der Merkmalsträger in allen Geschwisterreihen in die Berechnungen eingeht, ist der rechnerische Aufwand der Methode der Maximum Likelihood Scores bei Verwendung geeigneter Tabellen besonders gering.

Ohne auf die humangenetischen Folgerungen einzutreten, werden die verschiedenen Schätzverfahren vergleichend illustriert am Beispiel der rezessiven Taubstummheit (für die einfach abgeschnittene Binomialverteilung, $\rho = 1$) bzw. am Beispiel des Retinaglioms in Geschwisterreihen mit merkmalsfreien Eltern (für die doppelt abgeschnittene Binomialverteilung, $\rho = 2$). Es werden die Schätzungen samt Standardfehlern unter der Annahme einer einfachen ($k = 0$) bzw. vollständigen ($k = 1$) Auslese angegeben.

BIBLIOGRAPHIE

- APERT, E.: The laws of Naudin-Mendel. *J. Hered.* 5, 492-497, 1914.
- BAILEY, N.T.J.: The estimation of gene frequencies of recessives with incomplete multiple selection. *Ann. of Eugen.*, 16/3, 215-222, 1951.
- : A classification of methods of ascertainment and analysis in estimating the frequencies of recessives in man. *Ann. of Eugen.*, 16/3, 223-225, 1951.
- BERNSTEIN, F.: Zusammenfassende Betrachtungen über die erblichen Blutstrukturen des Menschen. *Z. Abstammungslehre*, 37, 237-270, 1925.
- : Ueber die Ermittlung und Prüfung von Genhypothesen aus Vererbungsbeobachtungen am Menschen und über die Unzulässigkeit der Weinbergischen Geschwistermethode als Korrektur der Auslesewirkung. *Arch. f. Rassenbiol.*, 22, 241-244, 1929.
- : Ueber die Unzulässigkeit der Weinbergischen Geschwistermethode. *Arch. f. Rassenbiol.*, 23, 285-290, 1930.
- : Ist die Weinbergische Geschwistermethode neben der direkten Methode zu benutzen? *Z. Abstammungslehre*, 58, 434, 1931.
- : Korrekturen bei erblichkeitsmathematischer Untersuchung von Krankheiten mit rezessivem Erbgang. *Arch. f. Rassenbiol.*, 27, 25-31, 1933.
- BERWALD, Fr. R.: Ueber die Weinbergische Geschwistermethode. *Skandinavisk Aktuarietidskrift*, 7, 75-106, 1924.
- FINNEY, D.J.: Systems of scores for estimating genetic parameters. *Proc. Eighth Internat. Congr. of Genetics*, Stockholm 1949.

- : Scores for the estimation of parameters. Second Internat. Biometric Conference, Geneva, 1949. - v.: Biometrics, 6, 221, 1950.
- : The truncated binomial distribution. Ann. of Eugen., London, 14, 319-328, 1947/49.
- FISHER, R.A.: On the mathematical foundation of theoretical statistics. Philos. Trans. Roy. Soc. London, A 222, 309-368, 1922.
- : The effect of methods of ascertainment upon the estimation of frequencies. Ann. of Eugen., 6, 13-25, 1934/35.
- HALDANE, J.B.S.: A method for investigating recessive characters in man. J. Genetics, 25, 251-255, 1932.
- : The estimation of the frequencies of recessive conditions in man. Ann. of Eugen., 8, 255-262, 1937/38.
- : A test for homogeneity of records of familial abnormalities. Ann. of Eugen., 14, 339-341, 1947/49.
- JUST, G.: Der Nachweis von Mendel-Zahlen bei Formen mit niedriger Nachkommenzahl. 1. Teil Arch. mikr. Anat., 94, 604-652, 1920.
- : Ueber die Ausschaltung des Recessivenüberschusses. Arch. f. Rassenbiol., 23, 260-275, 1931.
- : Weiteres über die Ausschaltung des Recessivenüberschusses nebst zwei Tabellen für den praktischen Gebrauch. Z. Abstammungslehre, 56, 112, 1930.
- KOLLER, S.: Methodik der menschlichen Erbforschung. II. Die Erbstatistik in der Familie. Handbuch der Erbbiologie des Menschen, II, 261-284, Berlin, J. Springer, 1940.
- LENZ, F.: Methoden der menschlichen Erblchkeitsforschung. In: Handbuch der hygienischen Untersuchungsmethoden, herausgegeben von E. Gotschlich. Dritter Band, 689-739, 1929.
- LUDWIG, W. und BOOST, Ch.: Vergleichende Wertung der Methoden zur Analyse rezessiver Erbgänge beim Menschen. Zeitschr. f. menschl. Vererbungs- und Konstitutionslehre, 24, 577-619, 1940.
- MOORE, P.G.: The estimation of the Poisson parameter from a truncated distribution. Biometrika, 39, 3/4, 247-251, 1952.
- : A note on truncated Poisson distributions. Biometrics, 10/3, 402-406, 1954.
- RAO, C.R.: Advanced statistical methods in biometric research. Wiley, New York, 1952.
- RIDER, P.R.: Truncated binomial and negative binomial distributions. Journ. of the American Statistical Assoc., 50, No 271, 877-883, 1955.
- VON BEHR, J.: Ueber die einfache und die allgemeine Geschwistermethode von Weinberg. Zeitschr. f. angewandte Math. und Mechanik, 7, 298-313, 1927.
- VON MISES, R.: Ueber die Weinbergsche «Geschwistermethode», Assekuranzjahrbuch Bd. 50, 40-52, Wien und Leipzig 1931.
- : Wahrscheinlichkeitsrechnung. Bd. 1 der Vorlesungen aus dem Gebiete der angewandten Mathematik. Fr. Deuticke, Leipzig und Wien 1931.
- WEINBERG, W.: Methoden und Fehlerquellen der Untersuchung auf Mendelsche Zahlen beim Menschen. Arch. Rassen- und Gesellschaftsbiologie, 9, 165-174, 1912.
- : Zur Vererbung der Anlage der Bluterkrankheit mit methodologischen Ergänzungen meiner Geschwistermethode. Arch. Rassen- und Gesellschaftsbiologie, 9, 694-709, 1912.
- : Auslesewirkungen bei biologisch-statistischen Problemen. Arch. Rassenbiologie, 10, 417-451, 557-581, 1913; 11, 569-573, 1914/15.
- : Mathematische Grundlagen der Probandenmethode. Zeitschr. ind. Abstammungs- und Vererbungslehre, 48, 179-228, 1928.
- : Zur Probandenmethode. Arch. f. Rassen- und Gesellschaftsbiologie, 23, 275-284, 417-421, 1931.

Adresse de l'auteur:

Alois KÄELIN, Zederstrasse 4, Zurich.

s	0.05	0.10	0.15	0.20	0.25	0.30	0.35	0.40	0.45	
2	11.073	6.156	4.583	3.858	3.483	3.295	3.229	3.255	3.363	
3	22.510	12.709	9.600	8.188	7.480	7.149	7.061	7.158	7.417	
4	34.310	19.654	15.038	12.967	11.948	11.489	11.386	11.545	11.925	
5	46.471	26.986	20.882	18.163	16.833	16.230	16.077	16.241	16.661	
6	58.992	34.697	27.113	23.739	22.071	21.279	21.008	21.090	21.448	
7	71.871	42.777	33.708	29.651	27.598	26.545	26.069	25.974	26.177	
8	85.105	51.216	40.641	35.856	33.347	31.947	31.172	30.814	30.795	s
9	98.692	60.002	47.885	42.308	39.258	37.416	36.257	35.568	35.287	2
10	112.628	69.121	55.412	48.962	45.274	42.898	41.282	40.220	39.665	3
11	126.909	78.558	63.191	55.775	51.350	48.356	46.228	44.772	43.948	4
12	141.533	88.300	71.193	62.706	57.445	53.762	51.088	49.235	48.157	5
13	156.494	98.330	79.389	69.721	63.531	59.103	55.865	53.623	52.312	6
14	171.788	108.630	87.749	76.789	69.585	64.372	60.566	57.952	56.429	7
15	187.411	119.186	96.247	83.881	75.592	69.568	65.203	62.235	60.519	8
16	203.358	129.978	104.857	90.976	81.543	74.697	69.785	66.485	64.592	9
s	0.95	0.90	0.85	0.80	0.75	0.70	0.65	0.60	0.55	10
2	38.191	18.365	11.861	8.681	6.827	5.635	4.824	4.252	3.843	11
3	62.716	32.465	22.250	17.072	13.932	11.830	10.339	9.246	8.434	12
4	84.171	44.289	31.028	24.398	20.409	17.740	15.830	14.408	13.330	13
5	105.260	55.531	39.134	31.060	26.301	23.189	21.011	19.416	18.217	14
6	126.316	66.663	47.041	37.445	31.867	28.300	25.880	24.174	22.944	15
7	147.368	77.777	54.898	43.735	37.288	33.222	30.531	28.711	27.478	16
8	168.421	88.889	62.744	49.996	42.652	38.051	35.055	33.093	31.844	
9	189.474	100.000	70.588	56.249	47.995	42.840	39.511	37.377	36.088	
10	210.526	111.111	78.431	62.500	53.332	47.613	43.935	41.605	40.250	
11	321.579	122.222	86.275	68.750	58.666	52.379	48.343	45.804	44.360	
12	252.632	133.333	94.118	75.000	64.000	57.142	52.743	49.986	48.439	
13	273.684	144.444	101.961	81.250	69.333	61.904	57.141	54.160	52.501	
14	294.737	155.556	109.804	87.500	74.667	66.667	61.538	58.330	56.553	
15	315.789	166.667	117.647	93.750	80.000	71.429	65.934	62.499	60.599	
16	336.842	177.778	125.490	100.000	85.333	76.190	70.330	66.666	64.643	

s	0.05	0.10	0.15	0.20	0.25	0.30	0.35	0.40	0.45	
3	7.510	4.252	3.228	2.774	2.560	2.480	2.493	2.583	2.749	
4	15.272	8.790	6.783	5.923	5.551	5.453	5.557	5.825	6.258	
5	23.290	13.623	10.676	9.459	8.984	8.935	9.188	9.698	10.455	
6	31.566	18.757	14.916	13.391	12.864	12.908	13.352	14.124	15.199	
7	40.105	24.196	19.508	17.721	17.182	17.345	17.986	18.993	20.309	
8	48.910	29.947	24.457	22.446	21.919	22.204	23.006	24.169	25.601	s
9	57.982	36.013	29.763	27.559	27.049	27.426	28.317	29.519	30.918	3
10	67.326	42.399	35.426	33.045	32.536	32.945	33.820	34.924	36.145	4
11	76.944	49.105	41.443	38.886	38.337	38.689	39.423	40.290	41.215	5
12	86.839	56.135	47.806	45.058	44.404	44.587	45.045	45.552	46.102	6
13	97.012	63.488	54.507	51.534	50.686	50.574	50.624	50.675	50.808	7
14	107.467	71.165	61.534	58.283	57.134	56.592	56.117	55.645	55.354	8
15	118.205	79.164	68.874	65.272	63.698	62.592	61.496	60.466	59.766	9
16	129.228	87.482	76.511	72.467	70.332	68.537	66.749	65.151	64.074	10
s	0.95	0.90	0.85	0.80	0.75	0.70	0.65	0.60	0.55	11
3	52.197	23.148	13.923	9.566	7.111	5.580	4.563	3.858	3.358	12
4	82.559	41.422	27.219	19.910	15.451	12.474	10.376	8.849	7.718	13
5	105.078	54.875	37.814	28.978	23.435	19.572	16.710	14.518	12.811	14
6	126.298	66.538	46.659	36.633	30.462	26.167	22.929	20.365	18.281	15
7	147.367	77.756	54.800	43.454	36.675	32.099	28.713	26.029	23.801	16
8	168.421	88.885	62.721	49.907	42.407	37.509	34.024	31.348	29.151	
9	189.474	99.999	70.583	56.222	47.903	42.594	38.963	36.311	34.230	
10	210.526	111.111	78.430	62.492	53.299	47.507	43.657	40.986	39.029	
11	231.579	122.222	86.274	68.748	58.655	52.355	48.208	45.458	43.590	
12	252.632	133.333	94.118	74.999	63.996	57.124	52.680	49.799	47.969	
13	273.684	144.444	101.961	81.250	69.332	61.898	57.112	54.062	52.222	
14	294.737	155.556	109.804	87.500	74.666	66.664	61.525	58.280	56.391	
15	315.789	166.667	117.647	93.750	80.000	71.428	65.928	62.473	60.508	
16	336.842	177.778	125.490	100.000	85.333	76.190	70.327	66.653	64.592	

Table 1 b: Méthode de Maximum Likelihood, Poids w_s , $\rho = 2$.

s	0.05	0.10	0.15	0.20	0.25	0.30	0.35	0.40	0.45	
2	21.039	11.080	7.792	6.173	5.224	4.614	4.198	3.906	3.700	s
3	21.016	11.029	7.707	6.047	5.049	4.379	3.891	3.514	3.205	2
4	20.984	10.958	7.589	5.875	4.815	4.073	3.506	3.041	2.639	3
5	20.942	10.868	7.442	5.664	4.533	3.717	3.075	2.540	2.075	4
6	20.892	10.758	7.266	5.417	4.214	3.331	2.631	2.053	1.568	5
7	20.832	10.631	7.065	5.142	3.872	2.934	2.200	1.613	1.144	6
8	20.764	10.486	6.841	4.845	3.517	2.544	1.802	1.236	0.811	7
9	20.687	10.324	6.597	4.532	3.160	2.173	1.449	0.926	0.560	8
10	20.602	10.147	6.337	4.211	2.810	1.831	1.146	0.680	0.379	9
11	20.507	9.956	6.063	3.887	2.476	1.525	0.893	0.491	0.251	10
12	20.405	9.751	5.780	3.566	2.162	1.255	0.686	0.350	0.164	11
13	20.294	9.534	5.489	3.251	1.872	1.022	0.521	0.246	0.106	12
14	20.175	9.307	5.195	2.947	1.609	0.825	0.392	0.171	0.068	13
15	20.049	9.069	4.899	2.658	1.373	0.660	0.292	0.118	0.043	14
16	19.914	8.824	4.605	2.384	1.164	0.524	0.216	0.080	0.027	15
s	0.95	0.90	0.85	0.80	0.75	0.70	0.65	0.60	0.55	16
2	3.819	3.673	3.558	3.472	3.413	3.381	3.377	3.401	3.459	0.196
3	0.428	0.812	1.155	1.463	1.741	1.996	2.235	2.465	2.697	0.118
4	0.038	0.144	0.306	0.514	0.756	1.025	1.313	1.618	1.938	0.070
5	0.003	0.023	0.072	0.160	0.294	0.475	0.704	0.980	1.301	0.041
6	-	0.003	0.015	0.046	0.106	0.204	0.352	0.558	0.826	0.024
7	-	-	0.003	0.013	0.036	0.083	0.167	0.302	0.501	0.014
8	-	-	0.001	0.003	0.012	0.033	0.077	0.157	0.293	0.008
9	-	-	-	0.001	0.004	0.012	0.034	0.080	0.167	
10	-	-	-	-	0.001	0.005	0.015	0.039	0.093	
11	-	-	-	-	-	0.002	0.006	0.019	0.050	
12	-	-	-	-	-	0.001	0.003	0.009	0.027	
13	-	-	-	-	-	-	0.001	0.004	0.014	
14	-	-	-	-	-	-	-	0.002	0.007	
15	-	-	-	-	-	-	-	0.001	0.004	
16	-	-	-	-	-	-	-	-	0.002	

Table I c: Méthode de Maximum Likelihood. Valeurs b_s . $\rho = 1$. Abrégé d'après Finney, D. J., *Annals of Eugenics* 14 (1947/49).

s	0.05	0.10	0.15	0.20	0.25	0.30	0.35	0.40	0.45	s
3	42.093	22.194	15.638	12.426	10.560	9.375	8.588	8.058	7.710	0.50
4	42.074	22.150	15.563	12.311	10.393	9.142	8.269	7.629	7.139	3
5	42.048	22.092	15.461	12.154	10.165	8.825	7.840	7.060	6.399	4
6	42.016	22.017	15.331	11.953	9.876	8.426	7.310	6.379	5.547	5
7	41.977	21.926	15.172	11.709	9.527	7.954	6.700	5.622	4.648	6
8	41.931	21.818	14.985	11.422	9.124	7.421	6.035	4.835	3.769	7
9	41.878	21.693	14.768	11.095	8.672	6.842	5.341	4.060	2.963	8
10	41.818	21.552	14.523	10.729	8.179	6.232	4.646	3.331	2.264	9
11	41.751	21.393	14.251	10.329	7.654	5.609	3.974	2.675	1.684	10
12	41.676	21.217	13.951	9.898	7.107	4.989	3.345	2.105	1.224	11
13	41.595	21.024	13.626	9.440	6.548	4.387	2.774	1.626	0.871	12
14	41.506	20.814	13.277	8.962	5.986	3.816	2.267	1.235	0.608	13
15	41.409	20.587	12.905	8.468	5.432	3.284	1.829	0.924	0.418	14
16	41.305	20.344	12.513	7.963	4.892	2.797	1.457	0.682	0.283	15
s	0.95	0.90	0.85	0.80	0.75	0.70	0.65	0.60	0.55	16
3	10.700	9.722	8.980	8.418	8.000	7.701	7.506	7.407	7.404	0.50
4	1.598	2.829	3.762	4.467	5.004	5.427	5.778	6.094	6.406	3
5	0.178	0.631	1.249	1.946	2.652	3.357	4.013	4.631	5.219	4
6	0.017	0.119	0.353	0.732	1.242	1.855	2.538	3.262	4.006	5
7	0.001	0.020	0.089	0.247	0.524	0.936	1.481	2.146	2.907	6
8	-	0.003	0.021	0.077	0.205	0.439	0.809	1.330	2.005	7
9	-	-	0.005	0.023	0.076	0.195	0.418	0.785	1.323	8
10	-	-	0.001	0.006	0.027	0.083	0.207	0.444	0.840	9
11	-	-	-	0.002	0.009	0.034	0.099	0.243	0.516	10
12	-	-	-	-	0.003	0.013	0.046	0.129	0.308	11
13	-	-	-	-	0.001	0.005	0.021	0.067	0.180	12
14	-	-	-	-	-	0.002	0.009	0.034	0.103	13
15	-	-	-	-	-	0.001	0.004	0.017	0.058	14
16	-	-	-	-	-	-	0.002	0.008	0.032	15

Table 1d: Méthode de Maximum Likelihood, Valeurs b_s , $\rho = 2$.

s	0.05	0.10	0.15	0.20	0.25	0.30	0.35	0.40	0.45	
2	11.073	6.156	4.583	3.858	3.483	3.295	3.229	3.255	3.363	
3	22.141	12.300	9.147	7.684	6.919	6.523	6.363	6.378	6.543	
4	33.218	18.462	13.737	11.548	10.403	9.808	9.563	9.574	9.802	
5	44.316	24.666	18.390	15.494	13.985	13.209	12.892	12.909	13.204	
6	55.445	30.930	23.129	19.547	17.695	16.749	16.367	16.390	16.746	
7	66.616	37.271	27.972	23.726	21.542	20.433	19.984	20.004	20.406	
8	77.838	43.701	32.930	28.037	25.530	24.255	23.731	23.732	24.160	
9	89.119	50.229	38.013	32.485	29.655	28.206	27.588	27.550	27.982	s
10	100.467	56.864	43.226	37.068	33.910	32.270	31.540	31.440	31.855	2
11	111.887	63.613	48.569	41.784	38.284	36.435	35.568	35.385	35.764	3
12	123.386	70.480	54.044	46.625	42.766	40.685	39.658	39.371	39.700	4
13	134.970	77.469	59.649	51.586	47.347	45.008	43.798	43.390	43.655	5
14	146.643	84.583	65.382	56.659	52.015	49.392	47.979	47.433	47.624	6
15	158.410	91.822	71.238	61.835	56.758	53.827	52.191	51.495	51.604	7
16	170.275	99.188	77.215	67.106	61.569	58.305	56.428	55.571	55.592	8
s	0.95	0.90	0.85	0.80	0.75	0.70	0.65	0.60	0.55	9
2	38.191	18.365	11.861	8.681	6.827	5.635	4.824	4.252	3.843	10
3	60.008	30.030	20.068	15.121	12.190	10.277	8.955	8.013	7.335	11
4	81.356	41.383	28.084	21.463	17.523	14.936	13.135	11.842	10.905	12
5	102.564	52.632	36.039	27.787	22.879	19.656	17.407	15.787	14.606	13
6	123.711	63.830	43.957	34.093	28.242	24.408	21.740	19.818	18.418	14
7	144.828	75.000	51.852	40.385	33.602	29.173	26.104	23.903	22.306	15
8	165.926	86.154	59.733	46.667	38.957	33.942	30.483	28.018	26.243	16
9	187.013	97.297	67.606	52.941	44.308	38.710	34.870	32.151	30.212	
10	208.092	108.434	75.472	59.211	49.655	43.479	39.260	36.294	34.200	
11	229.167	119.565	83.333	65.476	55.000	48.246	43.651	40.443	38.200	
12	250.237	130.693	91.192	71.739	60.343	53.012	48.044	44.595	42.209	
13	271.304	141.818	99.048	78.000	65.684	57.778	52.437	48.750	46.224	
14	292.369	152.941	106.902	84.259	71.024	62.543	56.831	52.907	50.242	
15	313.433	164.062	114.754	90.517	76.364	67.308	61.224	57.065	54.264	
16	334.495	175.182	122.602	96.774	81.702	72.072	65.619	61.225	58.288	

Table II a : Méthode de Weinberg. Poids w_s . $\rho = 1$.

s	0.05	0.10	0.15	0.20	0.25	0.30	0.35	0.40	0.45	
3	7.510	4.252	3.228	2.774	2.560	2.480	2.493	2.583	2.749	s 3 4 5 6 7 8 9 10 11 12 13 14 15 16
4	14.556	8.017	5.940	4.998	4.528	4.317	4.279	4.380	4.613	
5	21.226	11.462	8.383	6.997	6.315	6.016	5.972	6.131	6.480	
6	27.589	14.696	10.687	8.919	8.078	7.740	7.737	8.002	8.515	
7	33.700	17.794	12.934	10.842	9.891	9.559	9.640	10.048	10.760	
8	39.606	20.808	15.172	12.812	11.797	11.510	11.707	12.287	13.218	
9	45.344	23.778	17.437	14.859	13.820	13.610	13.950	14.719	15.876	
10	50.947	26.732	19.754	17.003	15.973	15.867	16.367	17.332	18.710	
11	56.442	29.695	22.139	19.257	18.267	18.285	18.953	20.112	21.697	
12	61.851	32.683	24.606	21.630	20.705	20.860	21.698	23.039	24.813	
13	67.195	35.712	27.167	24.129	23.289	23.587	24.588	26.095	28.036	
14	72.490	38.793	29.828	26.758	26.016	26.458	27.610	29.264	31.349	
15	77.751	41.935	32.596	29.519	28.885	29.465	30.752	32.530	34.737	
16	82.991	45.148	35.476	32.413	31.891	32.597	33.998	35.878	38.188	
s	0.95	0.90	0.85	0.80	0.75	0.70	0.65	0.60	0.55	
3	52.197	23.148	13.923	9.566	7.111	5.580	4.563	3.858	3.358	
4	73.703	34.088	21.144	14.862	11.237	8.935	7.384	6.302	5.532	
5	95.122	45.247	28.720	20.546	15.730	12.604	10.456	8.929	7.827	
6	116.416	56.429	36.446	26.474	20.521	16.594	13.844	11.849	10.377	
7	137.634	67.597	44.225	32.520	25.489	20.806	17.483	15.030	13.183	
8	158.806	78.750	52.022	38.624	30.557	25.161	21.301	18.418	16.211	
9	179.949	89.893	59.830	44.761	35.685	29.605	25.241	21.958	19.417	
10	201.073	101.029	67.645	50.920	40.852	34.110	29.265	25.611	22.762	
11	222.183	112.160	75.464	57.093	46.047	38.656	33.351	29.346	26.215	
12	243.283	123.288	83.287	63.279	51.262	43.235	37.481	33.143	29.749	
13	264.375	134.413	91.113	69.474	56.494	47.838	41.646	36.988	33.348	
14	285.461	145.535	98.941	75.676	61.739	52.460	45.839	40.871	36.999	
15	306.542	156.656	106.677	81.884	66.994	57.099	50.055	44.785	40.691	
16	327.620	167.776	114.602	88.096	72.258	61.752	54.289	48.724	44.418	

Table II b: Méthode de Weinberg. Poids w_s . $\rho = 2$.

s	0.05	0.10	0.15	0.20	0.25	0.30	0.35	0.40	0.45	
2	11.073	6.156	4.583	3.858	3.483	3.295	3.229	3.255	3.363	
3	22.141	12.300	9.147	7.684	6.919	6.523	6.363	6.378	6.543	
4	33.218	18.462	13.737	11.548	10.403	9.808	9.563	9.574	9.802	
5	44.316	24.666	18.390	15.494	13.985	13.209	12.892	12.909	13.204	
6	55.445	30.930	23.129	19.547	17.695	16.749	16.367	16.390	16.746	
7	66.616	37.271	27.972	23.726	21.542	20.433	19.984	20.004	20.406	
8	77.838	43.701	32.930	28.037	25.530	24.255	23.731	23.732	24.160	
9	89.119	50.229	38.013	32.485	29.655	28.206	27.588	27.550	27.982	
10	100.467	56.864	43.226	37.068	33.910	32.270	31.540	31.440	31.855	s
11	111.887	63.613	48.569	41.784	38.284	36.435	35.568	35.385	35.764	2
12	123.386	70.480	54.044	46.625	42.766	40.685	39.658	39.371	39.700	3
13	134.970	77.469	59.649	51.586	47.347	45.008	43.798	43.390	43.655	4
14	146.643	84.583	65.382	56.659	52.015	49.392	47.979	47.433	47.624	5
15	158.410	91.822	71.238	61.835	56.758	53.827	52.191	51.495	51.604	6
16	170.275	99.188	77.215	67.106	61.569	58.305	56.428	55.571	55.592	7
										8
s	0.95	0.90	0.85	0.80	0.75	0.70	0.65	0.60	0.55	9
2	38.191	18.365	11.861	8.681	6.827	5.635	4.824	4.252	3.843	10
3	60.008	30.030	20.068	15.121	12.190	10.277	8.955	8.013	7.335	11
4	81.356	41.383	28.084	21.463	17.523	14.936	13.135	11.842	10.905	12
5	102.564	52.632	36.039	27.787	22.879	19.656	17.407	15.787	14.606	13
6	123.711	63.830	43.957	34.093	28.242	24.408	21.740	19.818	18.418	14
7	144.828	75.000	51.852	40.385	33.602	29.173	26.104	23.903	22.306	15
8	165.926	86.154	59.733	46.667	38.957	33.942	30.483	28.018	26.243	16
9	187.013	97.297	67.606	52.941	44.308	38.710	34.870	32.151	30.212	
10	208.092	108.434	75.472	59.211	49.655	43.479	39.260	36.294	34.200	
11	229.167	119.565	83.333	65.476	55.000	48.246	43.651	40.443	38.200	
12	250.237	130.693	91.192	71.739	60.343	53.012	48.044	44.595	42.209	
13	271.304	141.818	99.048	78.000	65.684	57.778	52.437	48.750	46.224	
14	292.369	152.941	106.902	84.259	71.024	62.543	56.831	52.907	50.242	
15	313.433	164.062	114.754	90.517	76.364	67.308	61.224	57.065	54.264	
16	334.495	175.182	122.605	96.774	81.702	72.072	65.619	61.225	58.288	

Table III a: Méthode de Rider. Poids w_s . $\rho = 1$.

s	0.05	0.10	0.15	0.20	0.25	0.30	0.35	0.40	0.45	
3	7.510	4.252	3.228	2.274	2.560	2.480	2.493	2.583	2.749	
4	15.183	8.684	6.654	5.766	5.360	5.223	5.270	5.470	5.817	
5	23.020	13.294	10.274	8.968	8.384	8.202	8.294	8.608	9.133	
6	31.017	18.078	14.081	12.368	11.616	11.396	11.535	11.960	12.650	
7	39.175	23.035	18.071	15.958	15.044	14.788	14.970	15.497	16.336	
8	47.492	28.161	22.237	19.731	18.656	18.362	18.581	19.195	20.158	
9	55.969	33.454	26.575	23.678	22.442	22.104	22.347	23.029	24.090	
10	64.602	38.910	31.081	27.794	26.393	26.001	26.252	26.977	28.103	s
11	73.393	44.528	35.751	32.071	30.498	30.040	30.277	31.016	32.173	3
12	82.338	50.305	40.580	36.504	34.750	34.207	34.404	35.125	36.282	4
13	91.439	56.239	45.564	41.087	39.138	38.488	38.616	39.286	40.413	5
14	100.692	62.327	50.700	45.812	43.651	42.869	42.895	43.484	44.556	6
15	110.099	68.568	55.984	50.675	48.281	47.336	47.229	47.707	48.702	7
16	119.657	74.958	61.412	55.667	53.016	51.877	51.604	51.946	52.847	8
s	0.95	0.90	0.85	0.80	0.75	0.70	0.65	0.60	0.55	9
3	52.197	23.148	13.923	9.566	7.111	5.580	4.563	3.858	3.358	10
4	77.795	37.671	24.281	17.612	13.653	11.064	9.268	7.974	7.025	11
5	100.401	50.238	33.415	24.945	19.835	16.426	14.008	12.227	10.889	12
6	122.189	62.108	42.010	31.902	25.793	21.693	18.756	16.566	14.895	13
7	143.661	73.669	50.326	38.632	31.590	26.872	23.489	20.952	19.002	14
8	164.984	85.073	58.485	45.218	37.269	31.972	28.188	25.356	23.173	15
9	186.225	96.389	66.553	51.711	42.862	37.002	32.845	29.752	27.377	16
10	207.415	107.652	74.562	58.144	48.395	41.978	37.461	34.129	31.591	
11	228.573	118.879	82.533	64.536	53.885	46.911	42.039	38.482	35.802	
12	249.710	130.082	90.478	70.898	59.343	51.811	46.587	42.809	40.001	
13	270.830	141.267	98.403	77.240	64.779	56.688	51.110	47.115	44.186	
14	291.938	152.440	106.315	83.566	70.197	61.545	55.612	51.401	48.354	
15	313.038	163.603	114.215	89.880	75.602	66.388	60.099	55.670	52.508	
16	334.130	174.758	122.107	96.184	80.997	71.219	64.574	59.926	56.648	

Table III b: Méthode de Rider. Poids w_s . $\rho = 2$.

s	0.05	0.10	0.15	0.20	0.25	0.30	0.35	0.40	0.45	
2	11.073	6.156	4.583	3.858	3.483	3.295	3.229	3.255	3.363	
3	22.141	12.300	9.147	7.684	6.919	6.523	6.363	6.378	6.543	
4	34.120	19.299	14.478	12.163	10.863	10.091	9.648	9.445	9.447	
5	46.392	26.752	20.392	17.298	15.468	14.253	13.401	12.817	12.476	
6	58.909	34.504	26.681	22.888	20.585	18.937	17.634	16.583	15.789	
7	71.694	42.545	33.276	28.798	26.039	23.974	22.224	20.692	19.414	
8	84.758	50.882	40.156	34.962	31.707	29.204	27.015	25.034	23.304	
9	98.107	59.519	47.313	41.339	37.511	34.514	31.882	29.497	27.385	
10	111.745	68.455	54.732	47.896	43.398	39.834	36.743	33.995	31.578	s
11	125.674	77.685	62.392	54.599	49.328	45.127	41.556	38.475	35.823	0.50
12	139.894	87.203	70.274	61.415	55.269	50.371	46.305	42.914	40.078	2
13	154.406	96.999	78.352	68.312	61.199	55.559	50.991	47.305	44.319	3
14	169.211	107.064	86.601	75.261	67.100	60.688	55.620	51.652	48.537	4
15	184.308	117.387	94.996	82.237	72.960	65.762	60.201	55.962	52.729	5
16	199.696	127.955	103.513	89.217	78.773	70.785	64.744	60.241	56.897	6
s	0.95	0.90	0.85	0.80	0.75	0.70	0.65	0.60	0.55	7
2	38.191	18.365	11.861	8.681	6.827	5.635	4.824	4.252	3.843	8
3	60.008	30.030	20.068	15.121	12.190	10.277	8.955	8.013	7.335	9
4	78.342	38.673	25.665	19.326	15.659	13.339	11.798	10.759	10.071	10
5	95.728	46.421	30.428	22.763	18.438	15.800	14.148	13.139	12.587	11
6	112.351	53.585	34.742	25.857	20.961	18.083	16.396	15.497	15.173	12
7	128.273	60.268	38.742	28.767	23.404	20.374	18.726	18.004	17.969	13
8	143.548	66.548	42.515	31.585	25.862	22.766	21.230	20.743	21.033	14
9	158.221	72.487	46.128	34.374	28.394	25.317	23.958	23.748	24.374	15
10	172.340	78.142	49.634	37.185	31.045	28.063	26.934	27.022	27.967	16
11	185.945	83.562	53.081	40.057	33.848	31.025	30.163	30.549	31.771	
12	199.075	88.790	56.508	43.022	36.826	34.215	33.638	34.300	35.738	
13	211.766	93.866	59.946	46.108	39.996	37.638	37.344	38.236	39.821	
14	224.054	98.823	63.424	49.335	43.371	41.287	41.255	42.320	43.977	
15	235.968	103.691	66.967	52.720	46.956	45.154	45.344	46.513	48.176	
16	247.540	108.498	70.596	56.276	50.753	49.223	49.581	50.781	52.393	

Table IV a: Méthode de Moore. Poids w_S . $\rho = 1$.

s	0.05	0.10	0.15	0.20	0.25	0.30	0.35	0.40	0.45	s
3	7.510	4.252	3.228	2.774	2.560	2.480	2.493	2.583	2.749	3
4	15.183	8.684	6.654	5.766	5.360	5.223	5.270	5.470	5.817	4
5	23.264	13.558	10.551	9.246	8.649	8.431	8.459	8.674	9.057	5
6	31.509	18.675	14.781	13.153	12.446	12.200	12.210	12.375	12.655	6
7	39.921	24.005	19.286	17.411	16.678	16.474	16.474	16.570	16.636	7
8	48.513	29.559	24.065	22.000	21.294	21.171	21.209	21.160	20.933	8
9	57.293	35.350	29.132	26.920	26.267	26.218	26.236	26.007	25.439	9
10	66.267	41.387	34.497	32.174	31.572	31.547	31.461	30.976	30.042	10
11	75.437	47.679	40.171	37.761	37.181	37.091	36.782	35.958	34.652	11
12	84.809	54.233	46.157	43.674	43.061	42.785	42.122	40.883	39.212	12
13	94.387	61.055	52.459	49.168	48.566	48.379	47.422	45.710	43.697	13
14	104.173	68.152	59.077	56.412	55.456	54.379	52.644	50.429	48.100	14
15	114.172	75.528	66.008	63.193	61.877	60.178	57.770	55.043	52.428	15
16	124.386	83.189	73.248	70.211	68.384	65.925	62.791	59.563	56.693	16
s	0.95	0.90	0.85	0.80	0.75	0.70	0.65	0.60	0.55	s
3	52.197	23.148	13.923	9.566	7.111	5.580	4.563	3.858	3.358	3
4	77.795	37.671	24.281	17.612	13.653	11.064	9.268	7.974	7.025	4
5	95.694	46.296	30.164	22.314	17.757	14.835	12.840	11.420	10.382	5
6	112.349	53.570	34.694	25.745	20.742	17.696	15.753	14.841	13.634	6
7	128.273	60.267	38.734	28.740	23.336	20.222	18.417	17.420	16.926	7
8	143.548	66.547	42.514	31.578	25.840	22.707	21.084	20.414	20.345	8
9	158.221	72.487	46.127	34.372	28.388	25.295	23.890	23.565	23.930	9
10	172.340	78.142	49.634	37.184	31.043	28.054	26.902	26.922	27.687	10
11	185.945	83.562	53.081	40.057	33.847	31.022	30.148	30.496	31.599	11
12	199.075	88.790	56.508	43.022	36.825	34.214	33.632	34.271	35.635	12
13	211.766	93.866	59.946	46.108	39.996	37.637	37.341	38.222	39.760	13
14	224.054	98.823	63.424	49.335	43.371	41.287	41.254	42.312	43.942	14
15	235.968	103.691	66.967	52.720	46.956	45.154	45.344	46.509	48.156	15
16	247.540	108.498	70.596	56.276	50.753	49.223	49.580	50.779	52.381	16

Table IVb : Méthode de Moore. Poids w_s . $\rho = 2$.

$\begin{matrix} p \\ s \end{matrix}$	.05	.10	.15	.20	.25	.30	.35	.40	.45
	$\rho = 1$								
2	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
3	0.984	0.968	0.953	0.938	0.925	0.913	0.901	0.891	0.882
4	0.968	0.939	0.913	0.891	0.871	0.854	0.840	0.829	0.822
5	0.954	0.914	0.881	0.853	0.831	0.814	0.802	0.795	0.793
6	0.940	0.891	0.853	0.823	0.802	0.787	0.779	0.777	0.781
7	0.927	0.871	0.830	0.800	0.781	0.770	0.767	0.770	0.780
8	0.915	0.853	0.810	0.782	0.766	0.759	0.761	0.770	0.785
9	0.903	0.837	0.794	0.768	0.755	0.754	0.761	0.775	0.793
10	0.892	0.823	0.780	0.757	0.749	0.752	0.764	0.782	0.803
11	0.882	0.810	0.769	0.749	0.746	0.753	0.769	0.790	0.814
12	0.872	0.798	0.759	0.744	0.744	0.757	0.776	0.800	0.824
13	0.862	0.788	0.751	0.740	0.745	0.762	0.784	0.809	0.835
14	0.854	0.779	0.745	0.738	0.747	0.767	0.792	0.818	0.844
15	0.845	0.770	0.740	0.737	0.751	0.774	0.800	0.827	0.853
16	0.837	0.763	0.736	0.738	0.755	0.781	0.809	0.836	0.861
	$\rho = 2$								
3	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
4	0.953	0.912	0.876	0.844	0.816	0.791	0.770	0.752	0.737
5	0.911	0.841	0.785	0.740	0.703	0.673	0.650	0.632	0.620
6	0.874	0.784	0.716	0.666	0.628	0.600	0.579	0.567	0.560
7	0.840	0.735	0.663	0.612	0.576	0.551	0.536	0.529	0.530
8	0.810	0.695	0.620	0.571	0.538	0.518	0.509	0.508	0.516
9	0.782	0.660	0.586	0.539	0.511	0.496	0.493	0.499	0.513
10	0.757	0.630	0.558	0.515	0.491	0.482	0.484	0.496	0.518
11	0.734	0.605	0.534	0.495	0.476	0.473	0.481	0.499	0.526
12	0.712	0.582	0.515	0.480	0.466	0.468	0.482	0.506	0.538
13	0.693	0.562	0.498	0.468	0.459	0.466	0.486	0.515	0.552
14	0.675	0.545	0.485	0.459	0.455	0.468	0.492	0.526	0.566
15	0.658	0.530	0.473	0.452	0.453	0.471	0.500	0.538	0.581
16	0.642	0.516	0.464	0.447	0.453	0.476	0.509	0.551	0.596

Table V: Efficience relative de la méthode de Weinberg.

ρ s	.50	.55	.60	.65	.70	.75	.80	.85	.90	.95
					$\rho = 1$					
2	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
3	0.875	0.870	0.867	0.866	0.869	0.875	0.886	0.902	0.925	0.957
4	0.818	0.818	0.822	0.830	0.842	0.859	0.880	0.905	0.934	0.967
5	0.795	0.802	0.813	0.828	0.848	0.870	0.895	0.921	0.948	0.974
6	0.789	0.803	0.820	0.840	0.862	0.886	0.910	0.934	0.957	0.979
7	0.794	0.812	0.833	0.855	0.878	0.901	0.923	0.945	0.964	0.983
8	0.803	0.824	0.847	0.870	0.892	0.913	0.933	0.952	0.969	0.985
9	0.814	0.837	0.860	0.883	0.904	0.923	0.941	0.958	0.973	0.987
10	0.826	0.850	0.872	0.894	0.913	0.931	0.947	0.962	0.976	0.988
11	0.838	0.861	0.883	0.903	0.921	0.938	0.952	0.966	0.978	0.990
12	0.849	0.871	0.892	0.911	0.928	0.943	0.957	0.969	0.980	0.991
13	0.859	0.880	0.900	0.918	0.933	0.947	0.960	0.971	0.982	0.991
14	0.867	0.888	0.907	0.924	0.938	0.951	0.963	0.974	0.983	0.992
15	0.875	0.895	0.913	0.929	0.942	0.955	0.966	0.975	0.984	0.993
16	0.883	0.902	0.918	0.933	0.946	0.957	0.968	0.977	0.985	0.993
					$\rho = 2$					
3	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
4	0.725	0.717	0.712	0.712	0.716	0.727	0.746	0.777	0.823	0.893
5	0.613	0.611	0.615	0.626	0.644	0.671	0.709	0.759	0.825	0.905
6	0.561	0.568	0.582	0.604	0.634	0.674	0.723	0.781	0.848	0.922
7	0.538	0.554	0.577	0.609	0.648	0.695	0.748	0.807	0.869	0.934
8	0.532	0.556	0.588	0.626	0.671	0.721	0.774	0.829	0.886	0.943
9	0.537	0.567	0.605	0.648	0.695	0.745	0.796	0.848	0.899	0.950
10	0.547	0.583	0.625	0.670	0.718	0.766	0.815	0.862	0.909	0.955
11	0.561	0.601	0.646	0.692	0.739	0.785	0.830	0.875	0.918	0.959
12	0.577	0.620	0.666	0.711	0.757	0.801	0.844	0.885	0.925	0.963
13	0.594	0.639	0.684	0.729	0.773	0.815	0.855	0.894	0.931	0.966
14	0.611	0.656	0.701	0.745	0.787	0.827	0.865	0.901	0.936	0.969
15	0.627	0.672	0.717	0.759	0.799	0.837	0.873	0.907	0.940	0.971
16	0.642	0.688	0.731	0.772	0.810	0.847	0.881	0.913	0.944	0.973

Table V: Efficience relative de la méthode de Weinberg (suite).

$\rho \backslash s$	.05	.10	.15	.20	.25	.30	.35	.40	.45
				$\rho = 1$					
2	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
3	0.984	0.968	0.953	0.938	0.925	0.913	0.901	0.891	0.882
4	0.968	0.939	0.913	0.891	0.871	0.854	0.840	0.829	0.822
5	0.954	0.914	0.881	0.853	0.831	0.814	0.802	0.795	0.793
6	0.940	0.891	0.853	0.823	0.802	0.787	0.779	0.777	0.781
7	0.927	0.871	0.830	0.800	0.781	0.770	0.767	0.770	0.780
8	0.915	0.853	0.810	0.782	0.766	0.759	0.761	0.770	0.785
9	0.903	0.837	0.794	0.768	0.755	0.754	0.761	0.775	0.793
10	0.892	0.823	0.780	0.757	0.749	0.752	0.764	0.782	0.803
11	0.882	0.810	0.769	0.749	0.746	0.753	0.769	0.790	0.814
12	0.872	0.798	0.759	0.744	0.744	0.757	0.776	0.800	0.824
13	0.862	0.788	0.751	0.740	0.745	0.762	0.784	0.809	0.835
14	0.854	0.779	0.745	0.738	0.747	0.767	0.792	0.818	0.844
15	0.845	0.770	0.740	0.737	0.751	0.774	0.800	0.827	0.853
16	0.837	0.763	0.736	0.738	0.755	0.781	0.809	0.836	0.861
				$\rho = 2$					
3	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
4	0.994	0.988	0.981	0.974	0.966	0.957	0.948	0.939	0.929
5	0.988	0.976	0.962	0.948	0.933	0.918	0.903	0.888	0.873
6	0.983	0.964	0.944	0.924	0.903	0.883	0.864	0.847	0.832
7	0.977	0.952	0.926	0.901	0.876	0.853	0.832	0.816	0.804
8	0.971	0.940	0.909	0.879	0.851	0.827	0.808	0.794	0.787
9	0.965	0.929	0.893	0.859	0.830	0.806	0.789	0.780	0.779
10	0.960	0.918	0.877	0.841	0.811	0.789	0.776	0.772	0.777
11	0.954	0.907	0.863	0.825	0.796	0.776	0.768	0.770	0.781
12	0.948	0.896	0.849	0.810	0.783	0.767	0.764	0.771	0.787
13	0.943	0.886	0.836	0.797	0.772	0.761	0.763	0.775	0.795
14	0.937	0.876	0.824	0.786	0.764	0.758	0.764	0.781	0.805
15	0.931	0.866	0.813	0.776	0.758	0.756	0.768	0.789	0.815
16	0.926	0.857	0.803	0.768	0.754	0.757	0.773	0.797	0.825

Table VI: Efficience relative de la méthode de Rider.

p s	.50	.55	.60	.65	.70	.75	.80	.85	.90	.95
					$\rho = 1$					
2	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
3	0.875	0.870	0.867	0.866	0.869	0.875	0.886	0.902	0.925	0.957
4	0.818	0.818	0.822	0.830	0.842	0.859	0.880	0.905	0.934	0.967
5	0.795	0.802	0.813	0.828	0.848	0.870	0.895	0.921	0.948	0.974
6	0.789	0.803	0.820	0.840	0.862	0.886	0.910	0.934	0.957	0.979
7	0.794	0.812	0.833	0.855	0.878	0.901	0.923	0.945	0.964	0.983
8	0.803	0.824	0.847	0.870	0.892	0.913	0.933	0.952	0.969	0.985
9	0.814	0.837	0.860	0.883	0.904	0.923	0.941	0.958	0.973	0.987
10	0.826	0.850	0.872	0.894	0.913	0.931	0.947	0.962	0.976	0.988
11	0.838	0.861	0.883	0.903	0.921	0.938	0.952	0.966	0.978	0.990
12	0.849	0.871	0.892	0.911	0.928	0.943	0.957	0.969	0.980	0.991
13	0.859	0.880	0.900	0.918	0.933	0.947	0.960	0.971	0.982	0.991
14	0.867	0.888	0.907	0.924	0.938	0.951	0.963	0.974	0.983	0.992
15	0.875	0.895	0.913	0.929	0.942	0.955	0.966	0.975	0.984	0.993
16	0.883	0.902	0.918	0.933	0.946	0.957	0.968	0.977	0.985	0.993
					$\rho = 2$					
3	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
4	0.920	0.910	0.901	0.893	0.887	0.884	0.885	0.892	0.909	0.942
5	0.861	0.850	0.842	0.838	0.839	0.846	0.861	0.884	0.916	0.955
6	0.821	0.815	0.813	0.818	0.829	0.847	0.871	0.900	0.933	0.967
7	0.798	0.798	0.805	0.818	0.837	0.861	0.889	0.918	0.947	0.975
8	0.788	0.795	0.809	0.828	0.852	0.879	0.906	0.932	0.957	0.980
9	0.786	0.800	0.819	0.843	0.869	0.895	0.920	0.943	0.964	0.983
10	0.790	0.809	0.833	0.858	0.884	0.908	0.930	0.951	0.969	0.985
11	0.799	0.821	0.847	0.872	0.896	0.919	0.939	0.957	0.973	0.987
12	0.809	0.834	0.860	0.884	0.907	0.927	0.945	0.961	0.976	0.988
13	0.820	0.846	0.871	0.895	0.916	0.934	0.951	0.965	0.978	0.990
14	0.831	0.857	0.882	0.904	0.923	0.940	0.955	0.968	0.980	0.991
15	0.842	0.868	0.891	0.912	0.929	0.945	0.959	0.971	0.982	0.991
16	0.852	0.877	0.899	0.918	0.935	0.949	0.962	0.973	0.983	0.992

Table VI: Efficience relative de la méthode de Rider (suite).

$\begin{matrix} p \\ s \end{matrix}$	.05	.10	.15	.20	.25	.30	.35	.40	.45
				$\rho = 1$					
2	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
3	0.984	0.968	0.953	0.938	0.925	0.913	0.901	0.891	0.882
4	0.994	0.982	0.963	0.938	0.909	0.878	0.847	0.818	0.792
5	0.998	0.991	0.977	0.952	0.919	0.878	0.834	0.789	0.749
6	0.999	0.994	0.984	0.964	0.933	0.890	0.839	0.786	0.736
7	0.998	0.995	0.987	0.971	0.944	0.903	0.853	0.797	0.742
8	0.996	0.993	0.988	0.975	0.951	0.914	0.867	0.812	0.757
9	0.994	0.992	0.988	0.977	0.956	0.922	0.879	0.829	0.776
10	0.992	0.990	0.988	0.978	0.959	0.929	0.890	0.845	0.796
11	0.990	0.989	0.987	0.979	0.961	0.933	0.899	0.859	0.815
12	0.988	0.988	0.987	0.979	0.962	0.937	0.906	0.872	0.832
13	0.987	0.986	0.987	0.980	0.963	0.940	0.913	0.882	0.847
14	0.985	0.986	0.987	0.980	0.964	0.943	0.918	0.891	0.860
15	0.983	0.985	0.987	0.980	0.965	0.945	0.923	0.899	0.871
16	0.982	0.984	0.987	0.981	0.966	0.948	0.928	0.906	0.881
				$\rho = 2$					
3	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
4	0.994	0.988	0.981	0.974	0.966	0.957	0.948	0.939	0.929
5	0.999	0.995	0.988	0.978	0.963	0.944	0.921	0.894	0.866
6	0.998	0.996	0.991	0.982	0.968	0.945	0.914	0.876	0.833
7	0.995	0.992	0.989	0.983	0.971	0.950	0.917	0.872	0.819
8	0.992	0.987	0.984	0.980	0.972	0.953	0.922	0.875	0.818
9	0.988	0.982	0.979	0.977	0.971	0.956	0.927	0.881	0.823
10	0.984	0.976	0.974	0.974	0.970	0.958	0.930	0.887	0.831
11	0.980	0.971	0.969	0.971	0.970	0.959	0.933	0.892	0.841
12	0.977	0.966	0.966	0.969	0.970	0.960	0.935	0.897	0.851
13	0.973	0.962	0.962	0.968	0.970	0.960	0.937	0.902	0.860
14	0.969	0.958	0.960	0.968	0.971	0.961	0.938	0.906	0.869
15	0.966	0.954	0.958	0.968	0.971	0.961	0.939	0.910	0.877
16	0.963	0.951	0.957	0.969	0.972	0.962	0.941	0.914	0.885

Table VII: Efficience relative de la méthode de Moore.

s \ p	.50	.55	.60	.65	.70	.75	.80	.85	.90	.95
					$\rho = 1$					
2	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
3	0.875	0.870	0.867	0.866	0.869	0.875	0.886	0.902	0.925	0.957
4	0.771	0.756	0.747	0.745	0.752	0.767	0.792	0.827	0.873	0.931
5	0.715	0.691	0.677	0.673	0.681	0.701	0.733	0.778	0.836	0.909
6	0.693	0.661	0.641	0.634	0.639	0.658	0.691	0.739	0.804	0.889
7	0.693	0.654	0.627	0.613	0.613	0.628	0.658	0.706	0.775	0.870
8	0.705	0.660	0.627	0.606	0.598	0.606	0.632	0.678	0.749	0.852
9	0.723	0.675	0.635	0.606	0.591	0.592	0.611	0.653	0.725	0.835
10	0.745	0.695	0.649	0.613	0.589	0.582	0.595	0.633	0.703	0.819
11	0.767	0.716	0.667	0.624	0.592	0.577	0.583	0.615	0.684	0.803
12	0.787	0.738	0.686	0.638	0.599	0.575	0.574	0.600	0.666	0.788
13	0.806	0.758	0.706	0.654	0.608	0.577	0.567	0.588	0.650	0.774
14	0.823	0.778	0.726	0.670	0.619	0.581	0.564	0.578	0.635	0.760
15	0.837	0.795	0.744	0.688	0.632	0.587	0.562	0.569	0.622	0.747
16	0.850	0.810	0.762	0.705	0.646	0.595	0.563	0.563	0.610	0.735
					$\rho = 2$					
3	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
4	0.920	0.910	0.901	0.893	0.887	0.884	0.885	0.892	0.909	0.942
5	0.838	0.810	0.787	0.768	0.758	0.758	0.770	0.798	0.844	0.911
6	0.788	0.746	0.711	0.687	0.676	0.681	0.703	0.744	0.805	0.890
7	0.763	0.711	0.669	0.641	0.630	0.636	0.661	0.707	0.775	0.870
8	0.756	0.698	0.651	0.620	0.605	0.609	0.633	0.678	0.749	0.852
9	0.759	0.699	0.649	0.613	0.594	0.593	0.611	0.654	0.725	0.835
10	0.770	0.709	0.657	0.616	0.591	0.582	0.595	0.633	0.703	0.819
11	0.783	0.725	0.671	0.625	0.593	0.577	0.583	0.615	0.684	0.803
12	0.798	0.743	0.688	0.638	0.599	0.575	0.574	0.600	0.666	0.788
13	0.813	0.761	0.707	0.654	0.608	0.577	0.567	0.588	0.650	0.774
14	0.827	0.779	0.726	0.671	0.619	0.581	0.564	0.578	0.635	0.760
15	0.840	0.796	0.744	0.688	0.632	0.587	0.562	0.569	0.622	0.747
16	0.851	0.811	0.762	0.705	0.646	0.595	0.563	0.563	0.610	0.735

Table VII : Efficience relative de la méthode de Moore (suite).

$\begin{matrix} s \\ r \end{matrix}$	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	
1	8	17	12	14	7	5	9	1	2	2	1	2	1	1	-	
2	2	3	6	4	6	7	3	2	-	1	1	-	-	-	-	
3		1	2	8	-	1	4	-	3	1	1	-	-	-	1	
4			-	1	1	1	2	1	1	-	1	-	-	-	-	
5				1	-	-	1	-	-	-	-	-	-	-	-	
6					-	-	-	-	-	-	-	-	-	-	-	
7						-	-	-	-	-	-	-	1	-	-	
$\sum n_{sr}$	10	21	20	28	14	14	19	4	6	4	4	2	2	1	1	N = 150
$\sum m_{sr}$	12	26	30	55	23	26	40	9	15	7	10	2	8	1	3	R = 267
$\sum sn_{sr}$	20	63	80	140	84	98	152	36	60	44	48	26	28	15	16	S = 910

Table 1a : Surdi-mutité récessive, Répartition des sujets atteints par effectif de la fratrie. $s \geq 2$, $r \geq 1$.

$\begin{matrix} s \\ r \end{matrix}$	3	4	5	6	7	8	9	10	11	12	13	14	15	16
2	4	2	1	3	1	1	-	-	1	1	-	-	-	-
3	3	2	2	3	-	1	1	1	2	-	-	-	-	-
4		1	2	-	2	1	-	-	-	-	-	-	-	-
5			-	-	-	-	1	-	1	-	-	-	-	-
6				-	-	-	-	1	-	-	-	-	-	-
$\sum_r n_{sr}$	7	5	5	6	3	3	3	1	4	1	-	-	-	-
$\sum_r m_{sr}$	17	14	16	15	10	9	14	3	13	2	-	-	-	-
$\sum_r sn_{sr}$	21	20	25	36	21	24	27	10	44	12	-	-	-	-
$N = 38$ $R = 113$ $S = 240$														

Table 1b : Rétinoblastome. Fratries issues de parents d'apparence normale. Répartition des sujets atteints par effectif de la fratrie. $s \geq 3$, $r \geq 2$.

Suite : Table 4 : Méthode de Moore.
Rétinoblastome ($\rho = 2, k = 0$)

s	n_s	F_s	G_s	$\hat{p}_s$	$n_s w_s$ $p = 0.25$	$n_s w_s$ $p = 0.30$
3	7	6.00000	10.00000	0.60000	24.381	23.068
4	5	5.00000	9.00000	0.55556	34.595	32.616
5	5	4.33333	9.33333	0.46429	54.316	50.455
6	6	1.50000	7.50000	0.20000	92.810	85.519
7	3	1.50000	4.50000	0.33333	61.754	56.812
8	3	0.93333	3.93333	0.23729	78.116	71.921
9	3	2.33571	5.33571	0.43775	95.122	87.611
10	1	0.25000	1.25000	0.20000	37.511	34.514
11	4	1.01587	5.01587	0.20253	173.593	159.337
12	1	0	1.00000	0	49.328	45.127
	38	22.86825	56.86825	(0.40213)	701.526	646.969

$p' = 0.25: \hat{p}' = 201.049/701.526 = 0.28659;$ $\hat{p}' - p' = 0.03659$
 $p'' = 0.30: \hat{p}'' = 186.019/646.969 = 0.28752;$ $\hat{p}'' - p'' = -0.01248$
 Par interpolation linéaire: $\hat{p} = 0.28728, W(\hat{p}) = 660.848$
 $\hat{p} = 0.28728 \pm 0.03890$

Table 5: Méthode de Maximum Likelihood. Estimations individuelles $\hat{p}_s$.
Surdi-mutité récessive ($\rho = 1, k = 1$)

s	n_s	$\hat{p}_s$	$n_s w_s$ p = 0.20	$n_s w_s$ p = 0.25
2	10	0.33333	38.580	34.830
3	21	0.20653	171.955	157.078
4	20	0.26638	259.344	238.961
5	28	0.34577	508.568	471.316
6	14	0.20433	332.340	308.999
7	14	0.21776	415.116	386.373
8	19	0.23087	681.269	633.599
9	4	0.22470	169.232	157.031
10	6	0.23220	293.771	271.645
11	4	0.12022	223.098	205.398
12	4	0.19226	250.825	229.781
13	2	0	139.443	127.062
14	2	0.28300	153.577	139.171
15	1	0	83.881	75.592
16	1	0.17960	90.976	81.543
	150		3811.975	3518.378
p' = 0.20: $\hat{p}' = 849.978/3811.975 = 0.22298$;			$\hat{p}' - p' = 0.02298$	
p'' = 0.25: $\hat{p}'' = 785.487/3518.378 = 0.22325$;			$\hat{p}'' - p'' = -0.02675$	
Par interpolation linéaire:			$\hat{p} = 0.22310, W(\hat{p}) = 3676.333$	
			$\hat{p} = 0.22310 \pm 0.01649$	

Rétinoblastome ($\rho = 2, k = 1$)				
s	n_s	$\hat{p}_s$	$n_s w_s$ p = 0.30	$n_s w_s$ p = 0.35
3	7	0.69231	17.361	17.449
4	5	0.62884	27.280	27.785
5	5	0.59946	44.676	45.942
6	6	0.26838	77.446	80.112
7	3	0.43406	52.036	53.957
8	3	0.30905	66.612	69.010
9	3	0.51163	82.278	84.952
10	1	0.24025	32.945	33.820
11	4	0.25247	154.754	157.691
12	1	0	44.587	45.045
	38		599.975	615.763
p' = 0.30: $\hat{p}' = 208.995/599.975 = 0.34834$;			$\hat{p}' - p' = 0.04834$	
p'' = 0.35: $\hat{p}'' = 214.743/615.763 = 0.34874$;			$\hat{p}'' - p'' = -0.00126$	
Par interpolation linéaire:			$\hat{p} = 0.34873, W(\hat{p}) = 615.362$	
			$\hat{p} = 0.34873 \pm 0.04031$	

Suite : Table 5 : Méthode de Maximum Likelihood. Estimations individuelles $\hat{p}_s$.

Rétinoblastome ($\rho = 2, k = 0$)

s	n_s	$\hat{p}_s$	$n_s w_s$ $p = 0.25$	$n_s w_s$ $p = 0.30$
3	7	0.60000	24.381	23.068
4	5	0.54257	37.400	35.743
5	5	0.52106	59.740	57.445
6	6	0.20432	100.996	97.380
7	3	0.36288	66.214	63.836
8	3	0.24620	82.794	79.635
9	3	0.45475	100.042	95.840
10	1	0.18819	39.258	37.416
11	4	0.20120	181.097	171.594
12	1	0	51.350	48.356
	38		743.272	710.313
$p' = 0.25: \hat{p}' = 220.415/743.272 = 0.29655;$				
$p'' = 0.30: \hat{p}'' = 210.983/710.313 = 0.29703;$				
Par interpolation linéaire :				
$\hat{p} = 0.29700 \pm 0.03747$				

Table 6: Méthode des Scores selon Maximum Likelihood.

Surdi-mutité récessive ($\rho = 1, k = 1$)

s	n _s	r _s	W _s Y _s	W _s p = 0.20	Y _s	W _s Y _s	W _s p = 0.25	Y _s
2	10	12	13.270	38.580	0.34396	11.760	34.830	0.33764
3	21	26	35.513	171.955	0.20652	32.638	157.078	0.20778
4	20	30	70.000	259.344	0.26991	63.700	238.961	0.26657
5	28	55	185.158	508.568	0.36408	166.409	471.316	0.35307
6	14	23	67.912	332.340	0.20434	63.671	308.999	0.20606
7	14	26	90.512	415.116	0.21804	84.459	386.373	0.21859
8	19	40	157.945	681.269	0.23184	146.510	633.599	0.23123
9	4	9	38.122	169.232	0.22526	35.360	157.031	0.22518
10	6	15	68.484	293.771	0.23312	63.140	271.645	0.23244
11	4	7	28.202	223.098	0.12641	27.429	205.398	0.13354
12	4	10	48.236	250.825	0.19231	44.685	229.781	0.19447
13	2	2	5.998	139.443	0.04301	6.923	127.062	0.05449
14	2	8	44.106	153.577	0.28719	39.449	139.171	0.28346
15	1	1	3.592	83.881	0.04282	3.960	75.592	0.05239
16	1	3	16.366	90.976	0.17989	14.836	81.543	0.18194
150		267	873.416	3811.975	0.22912	804.929	3518.378	0.22878
p' = 0.20: $\hat{p}' = 873.416/3811.975 = 0.22912$;					$\hat{p}' - p' = 0.02912$			
p'' = 0.25: $\hat{p}'' = 804.929/3518.378 = 0.22878$;					$\hat{p}'' - p'' = -0.02122$			
Par interpolation linéaire: $\hat{p} = 0.22892$;					W($\hat{p}$) = 3642.158			
$\hat{p} = 0.22892 \pm 0.01657$								

Rétinoblastome ($\rho = 2, k = 1$)

s	n _s	r _s	W _s Y _s	W _s p = 0.35	Y _s	W _s Y _s	W _s p = 0.40	Y _s
3	7	17	14.609	17.449	0.83724	14.427	18.079	0.79800
4	5	14	20.193	27.785	0.72676	20.188	29.125	0.69315
5	5	16	31.130	45.942	0.67759	31.367	48.488	0.64690
6	6	15	22.074	80.112	0.27554	24.226	84.746	0.28587
7	3	10	23.856	53.957	0.44213	24.801	56.978	0.43527
8	3	9	21.455	69.010	0.31090	22.995	72.507	0.31714
9	3	14	45.515	84.952	0.53577	46.153	88.558	0.52116
10	1	3	8.541	33.820	0.25254	9.169	34.924	0.26254
11	4	13	41.247	157.691	0.26157	43.467	161.159	0.26972
12	1	2	5.446	45.045	0.12090	6.228	45.552	0.13672
38		113	234.066	615.763	0.38012	243.021	640.114	0.37965
p' = 0.35: $\hat{p}' = 234.066/615.763 = 0.38012$;					$\hat{p}' - p' = 0.03012$			
p'' = 0.40: $\hat{p}'' = 243.021/640.114 = 0.37965$;					$\hat{p}'' - p'' = -0.02035$			
Par interpolation linéaire: $\hat{p} = 0.37984$;					W($\hat{p}$) = 630.296			
$\hat{p} = 0.37984 \pm 0.03983$								

Table 7 : Tableau comparatif des estimations.

Méthode	Surd-mutité récessive $\rho = 1$		Rétinoblastome $\rho = 2$	
	$k = 1$	$k = 0$	$k = 1$	$k = 0$
Weinberg	0.24108 ± 0.01881 (0.01897)	0.15395 ± 0.01309 (0.01571)	0.32761 ± 0.05485 (0.05448)	0.28646 ± 0.04187 (0.04107)
Rider	0.24108 ± 0.01881 (0.01897)	0.15395 ± 0.01309 (0.01571)	0.33671 ± 0.04460 (0.04437)	0.28646 ± 0.04187 (0.04107)
Moore	0.22857 ± 0.01695 (0.01735)	0.15395 ± 0.01309 (0.01571)	0.34286 ± 0.04181 (0.04157)	0.28728 ± 0.03890 (0.03776)
Maximum Likelihood	0.22310 ± 0.01649	0.15395 ± 0.01309	0.34873 ± 0.04031	0.29700 ± 0.03747
$\hat{P}_S$ individuels	(0.01686)	(0.01571)	(0.04094)	(0.03668)
Maximum Likelihood	0.22892 ± 0.01657	0.15395 ± 0.01309	0.37984 ± 0.03983	0.31388 ± 0.03764
Scores	(0.01686)	(0.01571)	(0.04094)	(0.03668)

Entre parenthèses déviations standard pour la valeur $p = 0.25$.

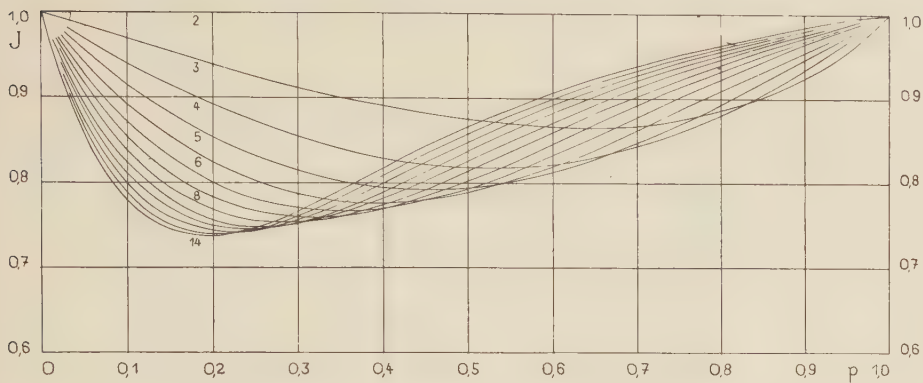


Figure 1 : Efficacité relative de la méthode d'estimation de Weinberg $\rho = 1$.

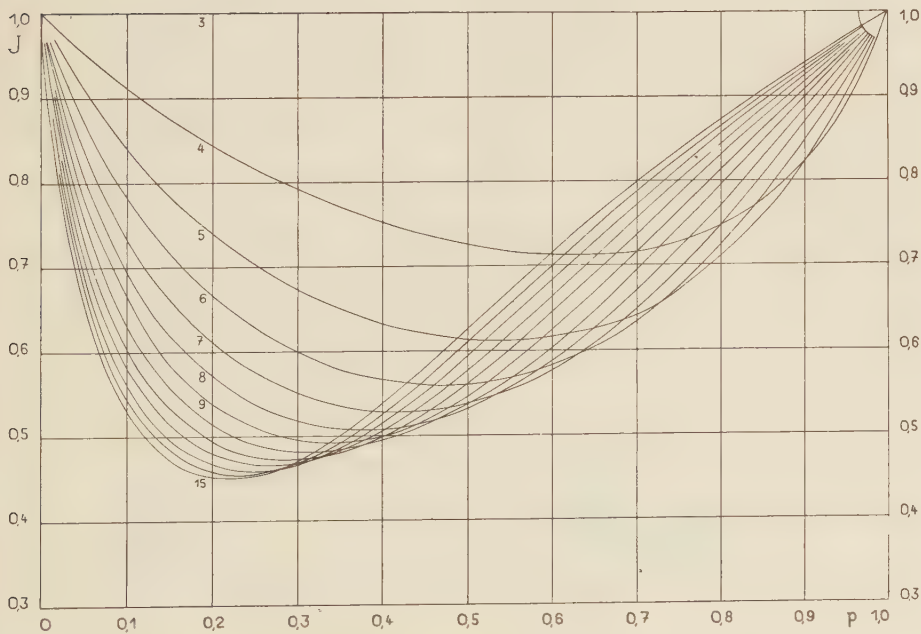


Figure 2 : Efficacité relative de la méthode d'estimation de Weinberg $\rho = 2$.

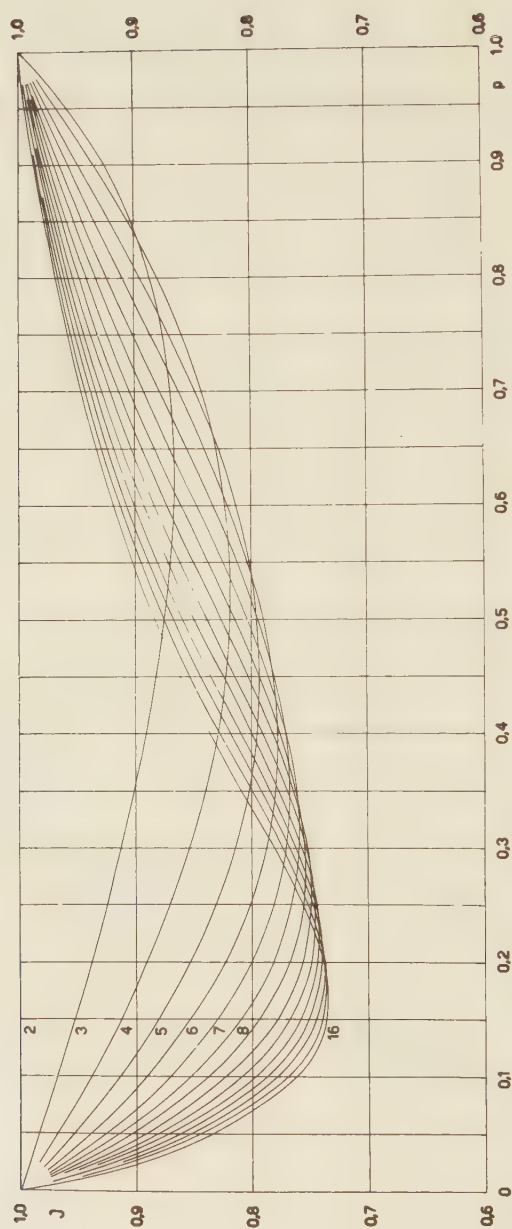


Figure 3 : Efficacité relative de la méthode d'estimation de Rider $\rho = 1$.

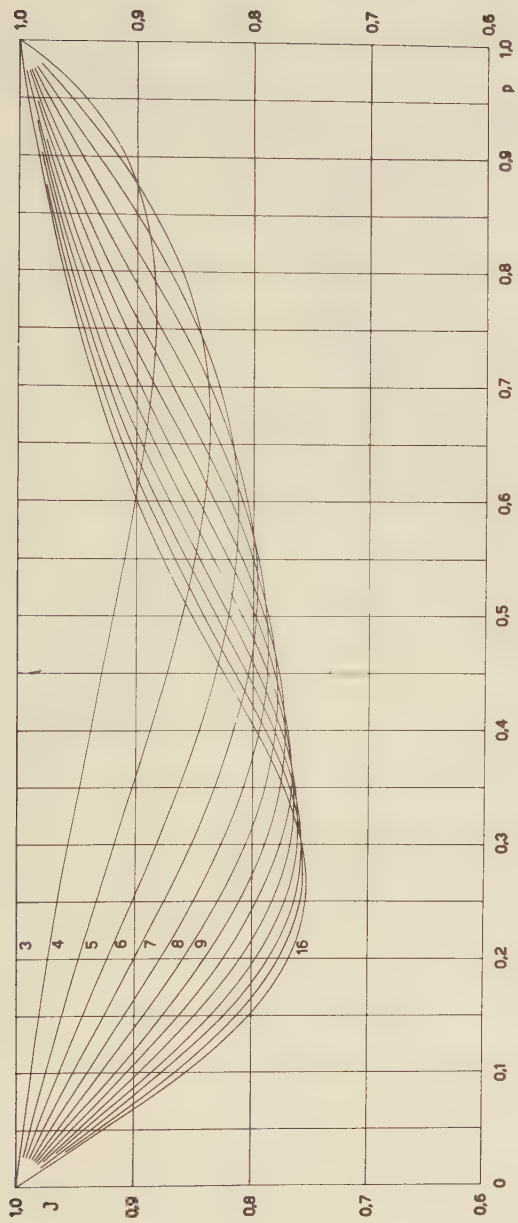


Figure 4 : Efficacité relative de la méthode d'estimation de Rider $\rho = 2$.

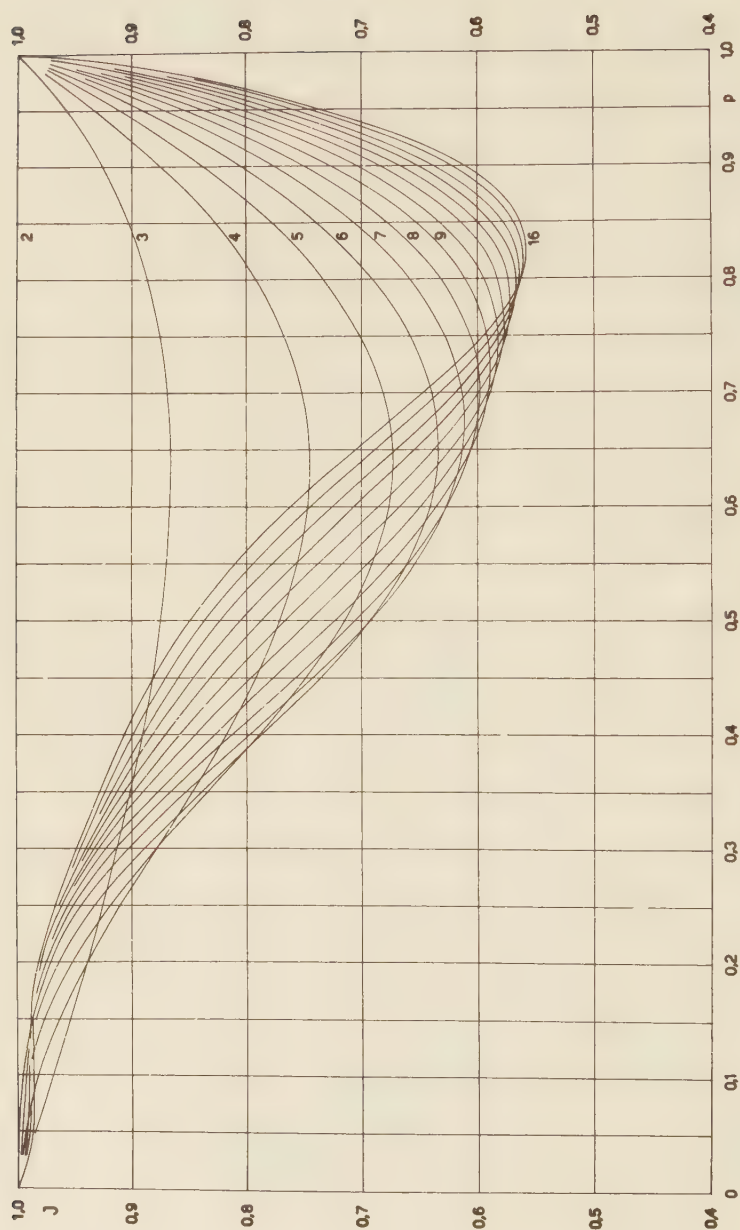


Figure 5 : Efficacité relative de la méthode d'estimation de Moore $\rho = 1$.

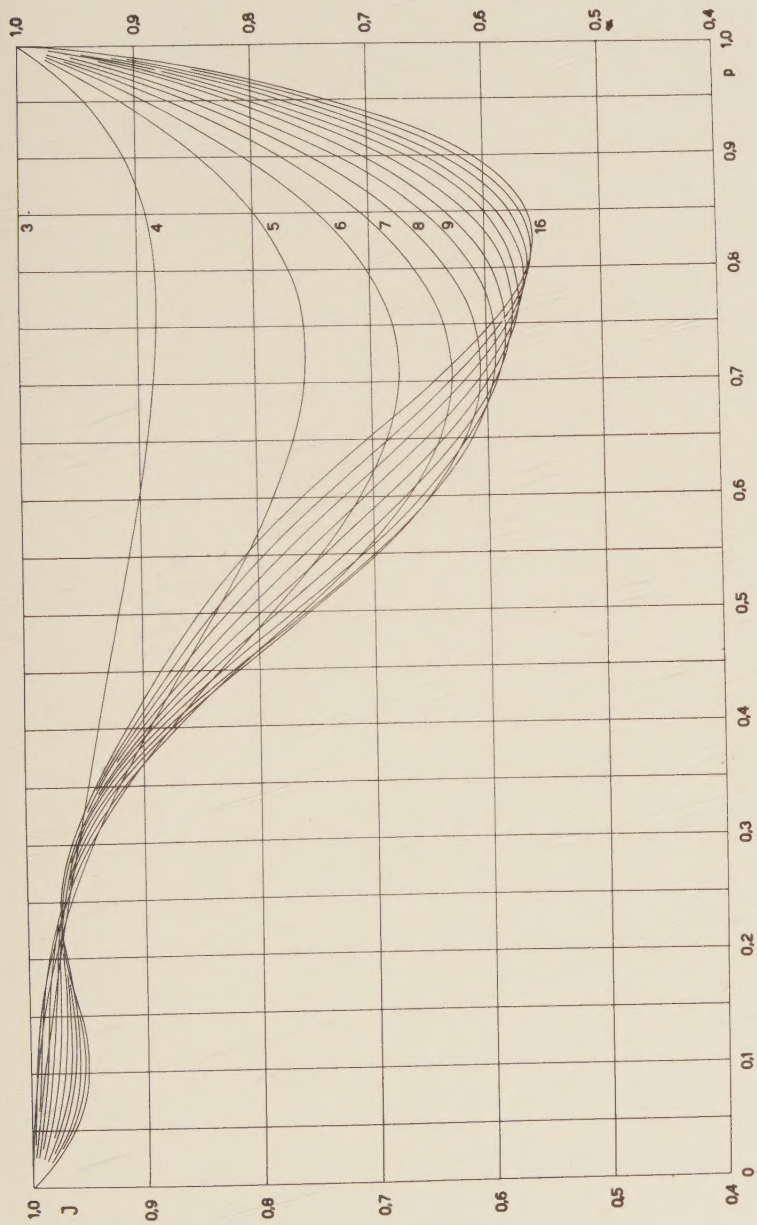


Figure 6 : Efficacité relative de la méthode d'estimation de Moore $\rho = 2$.

